

US muddles policy on fetal tissue

The failure of the US Congress to override the administration's veto of a bill to allow research with fetal tissue is a bad business, both for research and the civility of the coming election campaign.

THE United States, once the epitome of modernity, is coming to seem incorrigibly old-fashioned. The administration's decision that not a cent of federal money will go into fetal tissue research, and the failure last week of Congress to override President George Bush's veto of a bill that would have allowed it (see *Nature* **357**, 267; 1992), bears out the assertion. The administration's familiar and well-advertised objection is that most of the fetal tissue now available derives from voluntary abortions, so that encouragement for fetal tissue research would imply that the federal government condones voluntary abortion. (The argument that proof of therapeutic techniques based on the use of fetal tissue would even encourage voluntary abortion, fecklessly innocent as it is of the reasons why 1.5 million women a year in the United States seek abortions, is not to be taken seriously.) This position is not even ideological, but is symbolic only.

There are two kinds of potential uses for fetal tissue in the treatment of postnatal disease. Because it is immunologically immature, it is likely to be more readily compatible than mature tissue when used as grafts. And neuronal fetal tissue, the cells of which divide rapidly in embryogenesis, but hardly at all in adult life, may have a special part to play in neurological diseases such as parkinsonism (but that point has not yet been proved). Nobody would claim, at least at this stage, that the promise of the therapies that may follow from research with fetal tissue is comparable with that of gene therapy, for example. But research may show that some congenital diseases of the blood-forming system, as well as parkinsonism and other like conditions, may be tractable by fetal grafts and not easily by other means.

Morally, of course, it is irrelevant that fetal tissue grafting is likely never to be simple or a dominant part of medical practice: the mythical housemaid was not excused the illegitimacy of her baby because it was only a small one. Bush's White House thus argues that it stands on a matter of principle: if research shows fetal tissue to be valuable in medicine, and if aborted tissue is then used as grafts, the administration will be held to have connived in proving that aborted tissue has medical uses and will thus have condoned abortion. It believes its case is strengthened by the likelihood that aborted tissue would be used in the research. But, even in the White House view, that argument can be turned around. Would not the supposed sin of abortion be mitigated — not absolutely, but to some degree — by the use of fetal tissue in therapy?

That argument is not nearly as unprincipled as many theologians hold. The United States boasts of its ability to

accommodate diverse opinions. Now, there is a disturbing and sharp conflict on whether abortion should be legally permitted. (The Supreme Court's decision two weeks ago that states are free to act independently will make the conflict more audible by multiplying by fifty the number of campaigns for legislative change.) No wonder that abortion is likely to be prominent in the election campaign ahead.

But is it not the proper function of the government in a pluralistic society so to manage its affairs that substantial sections of opinion are not given a sense of being oppressed? That, indeed, is the line the White House has taken in its refusal to abridge the right of US citizens to buy guns (although Kalashnikovs and the like are now off bounds). Elsewhere, the abortion issue (nowhere settled) has been made manageable by subtlety blended with tolerance. To acknowledge that those who hold to the 'right to life' as well as to 'women's rights to self-determination' do so with great passion is not necessarily to allow that great passions make good policy. When some hold that voluntary abortion is deliberate murder, and others that it is a necessary instrument of social policy, serious conflict is unavoidable. People employ politicians to manage their affairs because they believe them to be skilled at the art of compromise. On this issue, the Bush administration has seemed more anxious to keep old friends than to win new ones.

Bush need not have dug in his heels on fetal tissue research. By doing so, he has given a hostage to fortune and an illustration of the administration's willingness to interfere with the administration of research. (The device of specifying what granting agencies may and may not do with federal funds is becoming uncomfortably familiar.) But Bush may also have delayed therapy for many who need it. □

Protest by barricade

Road-blocks are in the best French traditions, but this time deepen the shadow over the European enterprise.

THE book *Clochemerle*, later a successful film, which demonstrated the depths of Gallic passion even over small issues, is one recollection evoked by the barricades thrown over major highways last week by French truck drivers protesting at new strict regulations of drivers' licences. Another is the long period of student protest which, in 1968, made parts of Paris impassable — and which (among other



things) forced the government of France to re-lay the foundations of higher education, so giving France, with remarkable speed, its now-distinct flavour of modernity. But even though Bastille Day is only a few days away, and although barricades have been an honourable way of making public protests since 1789, the question must arise of whether the truck-drivers are onto something.

It is not as though it were just truck-drivers. Since the announcement a month ago of the amendment of the Common Agricultural Policy by the European Communities (EC), French farmers have been busily building barricades wherever the police would let them — usually of farm vehicles, but sometimes of burning tyres as at Lille at the weekend. The ironical culmination of this excitement is that the railway line from Lyon to Marseilles was blockaded both by farmers protesting at the new EC policy and by others incensed by the truck-drivers' blockage of the highways, which prevented them getting their produce to market. What does it all mean?

France is both conservative and, ordinarily, law-abiding. In the tradition of the barricade, the lawlessness is a challenge to the government. The issues the two groups raise are very different: the truck-drivers object to a scheme under which they would lose their driving licences automatically as they accumulate driving misdemeanours, the farmers object to the reduction of their subsidized prosperity. The passion is easily understood, but the government's position is also reasonable: France has a poor record on road safety, while France alone could not subsidize French farmers to the degree that the EC now recognizes to be unaffordable and impolitic.

What is to be done? Within the EC, France has been for 20 years one of the most solid supporters of European cohesion and one of the most protectionist of member states. It remains the first, on grand strategic grounds, but the end of Mme Edith Cresson's short spell as prime minister earlier in the year spelled the end of the second view. In reality, there has not been much time for the government to explain its change of tack, but it has also been less than energetic in the task. Now, the time is almost exhausted; inevitably, the farmers will influence the result of the referendum on the European Maastricht Treaty due in mid-September, not to mention the assembly elections due next year. Although the reform of agricultural policy has nothing to do with Maastricht, the government of France may yet regret not having explained more clearly what it is.

The truck-drivers, many of whom now cross European frontiers habitually, constitute a different but more easily soluble problem. They cannot reasonably object to stricter rules on road safety, but they can ask to be dealt with on the same footing as other European truck-drivers. The European Commission, which has already legislated on the numbers of hours a day a commercial driver can be at the wheel, might reasonably be asked to get France off the hook by decreeing uniform standards of driving for licence-holders. The trouble, there, is that Britain and like-minded devolutionary countries are emphasizing the virtues of the devolution called 'subsidiarity' to persuade their people to ratify Maastricht □

Keeping zoos alive

The London Zoological Society should either run its zoo with enthusiasm or hand over to others who would.

THE troubles of London's city-centre zoo are in danger of becoming boring. Every so often, the word is that the zoo is about to close, but then there is a last-minute rescue (most recently by the Emir of Kuwait. The danger is that people will become inured to the threat of catastrophe and that, somewhere along the line, there will be no rescue. That is why this is precisely the point at which the zoo's management must lay the foundations for the long-term survival of the London Zoo. The letter from Sir Barry Cross on page 102 gives no hint of how the Zoological Society of London will exert itself in that direction (which does not imply that nothing will be done).

Constitutionally and in practice, the London Zoo is a subsidiary of the Zoological Society, both of which have charitable status. Cross is right to say that the society will still have useful work to do even if the zoo is closed; whether its reputation (even among zoologists) would remain untarnished if that happened is another matter. And given that the society benefited both in reputation and in the scale on which its scientific work could be mounted during the long years of modest prosperity at the zoo, is it permissible now to contemplate walking away from a seemingly intractable problem? The society evidently has a sense that it is a dog being wagged by its tail, but also by the British government. But Cross's statement that the chief executive at the zoo in the past few years was "chosen" by the government does not absolve the society's council from responsibility.

What is now to be done? This journal holds that the London Zoo should be kept alive (and modernized) if at all possible — and that the outlook is not nearly as gloomy as supposed. What the zoo needs, first of all, is a membership scheme that offers ordinary people a sense of being in touch with modern zoology (biodiversity and all that) in return for an annual subscription. (That device brings the excellent San Diego Zoo more than half its revenue, but the London Zoo has a much greater catchment population.)

Although the Victorian Zoological Society was forever hearing first-hand accounts of Africa or the Amazon from just-returned explorers, London is probably even fuller now of people with that cast of mind, and with more exciting tales to tell. Could they not help to make membership worthwhile? Imaginative ways (such as shuttle buses) of getting people to the zoo (which, however central, is relatively inaccessible) would also help. But the most urgent need is that policy for the zoo should be determined by people with enthusiasm for its survival. Are the zoologists now so embarrassed by their zoo that they would let their subsidiary float off on its own? It is unthinkable that there is no group of well-wishers ready to give the London Zoo a further lease of life. Otherwise, the society has no choice but to muster the enthusiasm from among its own members. □

Controversial NIH genome researcher leaves for new \$70-million institute

Washington. The US National Institutes of Health (NIH) researcher whose patent applications for thousands of partial gene sequences have inflamed the genome community is leaving NIH to start what is expected to be the world's largest gene sequencing laboratory. J. Craig Venter announced this week that he will resign from NIH on 10 July to head a new \$70-million nonprofit research centre called the Institute for Genomic Research (IGR). By the time the privately funded institute gets up to speed, it is expected to include some 70 scientists and to be identifying some 2,000 gene markers a week.

Venter says the new institute aims to "do the genome project", starting from industrial-scale cDNA sequencing and moving to gene mapping and biology as well. With a core operation based around 18–20 robotic gene sequencers from Applied Biosystems Inc., Venter estimates that the institute will eventually be able to sequence as many as 60 million bases — one-fiftieth of the entire human genome — a year.

The IGR's initial funding comes from a ten-year grant from the Healthcare Investment Corporation, a venture capital group based in Edison, New Jersey, and backed by several corporate pension accounts. It has started other NIH-related biotechnology companies, including Genetic Therapy Inc. In addition to the IGR, the venture fund is creating a new biotechnology company, Human Genome Science Inc., to develop and market products based on IGR technology and discoveries. Both the new company and the IGR will be located in Montgomery County, Maryland, near NIH.

Venter says the IGR will initially concentrate on a ten-fold scale-up of his efforts to sequence random fragments of cDNA, known as Expressed Sequence Tags (ESTs). Researchers hope that a database of ESTs, which are markers for expressed genes, will help them to link genes with known function in model species (such as yeast and the mouse) to genes that may have similar function in humans. There are an estimated 50,000–100,000 human genes, for which fewer than 10,000 ESTs have been sequenced.

IGR researchers will also work on full sequencing of DNA clones, with an eventual target of sequencing an entire chromosome. Other projects will include sequencing full cDNAs (rather than just fragments), and the sequencing of the genomes of model species such as yeast, *Caenorhabditis elegans* and *Drosophila*, Venter says. A parallel effort will focus on genome biology, including research on gene expression and receptors.

Venter intends to take his NIH team with

him, and will soon be joined by his wife, Claire Fraser, a molecular receptor biology researcher, and her team at the National Institute on Alcohol Abuse and Alcoholism. Other recruits are expected to include specialists in gene sequencing technology and a "major gene mapper", Venter says. He is negotiating with IBM for an additional grant, on the order of several tens of million dollars, to fund cooperative development of genome-related computer technology.

The new institute appears to accomplish much of what Frederick Bourke, a wealthy US entrepreneur, had hoped to create last year (see *Nature* 355, 483; 1992). After Bourke's international recruiting effort failed to entice several key researchers, his project lost momentum and now appears to be in abeyance. Unlike Bourke's proposed company, IGR is not intended to be supported by government grants or outside contracts; it may, in fact, send money the other way, in the form of cooperative agreements to support research at NIH.

Given the controversy about Venter's gene patents at NIH, the intellectual property policy of the new institute is expected to be watched closely. Wallace Steinberg, president of the Healthcare Investment Corporation, says that the founders have not yet settled on a patent policy, "but we will be studying the problem very carefully". Other companies started by the venture capital fund have adopted the standard industry

policy of patenting genes only when they are fully sequenced and their function is known, and the new company is expected to start out with that position. Its final decision will depend on both the policy of the NIH and any decisions on the issue by the US patent office, Steinberg says.

News of the new institute was welcomed by several researchers this week as both an acknowledgement that the genome project is ready for scale-up and that industry recognizes its long-term commercial potential. "I think that everyone agrees that they don't want to let [large-scale] sequencing choke the academic community," says Richard Gibbs, a sequencing research at the Baylor College of Medicine. "I think the community is going to very pleased" to see industry fund the labour and technology intensive sequencing effort, leaving much of the mapping and gene biology to smaller university research teams, he says. "I can't see how this is anything but a shot in the arm for the whole [genome] project."

Victor McKusick, the Johns Hopkins University researcher who pioneered the genome project, predicts that there will be an "initial sense of unease" at the prospect of the world's largest gene sequencing operation being run by as controversial a figure as Venter. But as a whole, "the trajectory [of the initiative] is in the right direction. This is what it's going to take to get the job done."

Christopher Anderson

Solar observatory shaken by quake

Washington. Among the casualties of last week's double earthquake in the sparsely populated Mojave desert of California were telescopes at the Big Bear Solar Observatory, about 80 miles northeast of Los Angeles. The epicentre of the second earthquake, which measured 6.5 on the Richter scale, was a mere five miles from the observatory, which is on a small island in Big Bear Lake. Damage to the observatory itself was amplified because its foundation rests on soft lake-bottom soil; the main office on the shore was damaged only to the extent of a few cracked walls.

The concrete pedestal and mounting of the largest telescope, a 26-inch reflector, was destroyed and will have to be rebuilt. The two other main instruments, an eight-inch and ten-inch refractor, were also put out of action because of damage to their mountings. The good news, according to Bill Marquette, chief observer at Big Bear,

is that the optical equipment of all three telescopes appears to have escaped damage.

Marquette estimates that it will take a month or two, and about \$300,000, to rebuild the mountings, which had been damaged and then strengthened after a 4.9-magnitude earthquake in 1979. The California Institute of Technology, which owns and operates the observatory, has promised to help pay for the reconstruction, and additional support may come from one or more of the federal agencies that support research there.

The period of inaction will mean a considerable loss of data. Big Bear normally produces two or three thousand solar observations every day, and about 3,000 feet of 35-mm film a week, and distributes it worldwide. No other solar observatory produces such a quantity or range of fundamental solar data.

David Lindley

Search for contaminant in EMS outbreak goes slowly

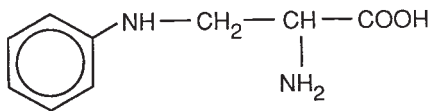
Tokyo & Washington. Hopes for a swift and conclusive investigation into the causes of an outbreak in the United States of a fatal blood disorder have foundered on the difficult nature of the problem facing researchers in Japan and the United States.

The outbreak of eosinophilia-myalgia syndrome (EMS), in which at least 27 people died and more than 1,500 were affected in 1989 and 1990, has been traced to consumption of a dietary supplement, L-tryptophan, that is made with the help of a genetically engineered organism. That circumstance has led to questions about the safety of the bio-engineering process itself. The case has also drawn unwanted public attention to the manufacturer, Showa Denko, which has given grants of more than \$3 million to 22 research groups at universities and institutes in the United States, as well as supporting a smaller research effort in Japan.

Progress has nevertheless been slow. Scientists believe it could take several more years to find an answer. And they are not willing to speculate on whether the problem lies in the production process. L-tryptophan is an essential amino acid sold in tablet, capsule and powder form as a food supple-

ment. It is metabolized to serotonin, a neurotransmitter, and has been used to treat insomnia, depression and other disorders.

Contrary to initial expectations, it now appears that EMS "is not due to a single cause or the effect of a single contaminant", says Sam Page at the US Food and Drug



The latest candidate cause of EMS

Administration. "We're looking at an enormously complex inflammatory disease process." There are some 60 contaminants in L-tryptophan, and even pure L-tryptophan appears to cause EMS-like symptoms in some people.

In 1990, when scientists at Minnesota Department of Health announced that batches of suspect L-tryptophan associated

with the EMS outbreak were made with the help of a genetically engineered strain of *Bacillus amyloliquefaciens*, opponents of recombinant-DNA technology used the discovery to question the safety of all genetically engineered products (see *Nature* **346**, 787; 1990). A cloud has been hanging over the biotechnology industry ever since.

Although there is still no answer to the question of whether genetic engineering played a role in the epidemic, the delay in answering that question appears to be due more to the pace of science than to a wish to suppress the truth. Contrary to the allegations of some activists and EMS victims, the reason there is still no known cause of the outbreak "is more a technical problem than a coverup," says Vincent Falanga, a University of Miami researcher who is investigating L-tryptophan's molecular mechanisms. Most researchers say that Showa Denko has been cooperative in supplying data on the production processes and samples of the material, and scientists funded by the company say they have been free to publish their findings. Others point out, however, that most of the research funded by Showa Denko concerns basic biomedical questions about L-tryptophan rather than the exact cause of the contaminations.

For the past two years, researchers have focused on one contaminant (peak E) found in liquid chromatographs of the batches of L-tryptophan associated with EMS. But extensive animal tests by researchers at the US National Institutes of Health have failed to establish more than a partial link between this contaminant and EMS.

Last month, a group of US scientists and another group from Japan independently announced the chemical structure of a second contaminant in the suspect L-tryptophan that they think may be the cause of the EMS outbreak. The new contaminant was found by researchers at the Mayo Clinic in the United States and by another group at the National Institute of Hygienic Sciences in Tokyo and has been identified as 3-phenylamino-L-aniline. This chemical is very similar to the contaminant in cooking oil that caused an outbreak of EMS in Spain in 1981.

But further animal experiments are needed to determine if the new contaminant is the cause of EMS, and nothing is known about the pathway of formation of the new contaminant in the process of production of L-tryptophan. However, this has not stopped some Japanese newspapers from speculating that the contaminant is directly linked to the use of a genetically engineered organism.

Mitsuru Uchiyama, director-general of the National Institute of Hygienic Sciences and leader of the Japanese group that isolated the new contaminant, hopes to collaborate with Page in establishing the pathway of formation of the new contaminant. But researchers in both countries expect that it will take several more years to pin down the cause of EMS.

David Swinbanks & Christopher Anderson

Legal battles amid the science

Washington. US government scientists trying to track down the cause of EMS must deal with more than just scientific issues. Attorneys for Showa Denko, the Japanese company that produced the L-tryptophan implicated in an outbreak of the disease, have gone to great lengths to obtain the most recent results of government research into the disease. Some researchers say that their behaviour has had a chilling effect on independent investigations.

Their aggressive legal tactics are understandable, given that Showa Denko has been sued for more than a billion dollars by EMS sufferers. But to Esther Sternberg, an EMS researcher at the US National Institute of Mental Health (NIMH) who was the focus of one such skirmish, the company's efforts to acquire her pre-publication data crossed the line from legitimate inquiry to intimidation.

In late 1990, attorneys from Showa Denko's US law firm, Cleary, Gottlieb, Steen and Hamilton, filed several requests under the Freedom of Information Act (FOIA) for Sternberg's pre-publication data on EMS and L-tryptophan. Sternberg supplied some of the data on which her published work was based, but she balked at a request for her glass pathology slides and protocols of planned experiments. After reviewing the case, officials from NIMH and other federal agencies that were the subject of similar FOIA requests decided that government scientists were not required to make their data available before publication. So the Showa Denko lawyers wrote to Sternberg herself. On the advice of her attorney, she refused to respond. Company lawyers also called Sternberg's boss, the director of the NIMH, and tried unsuccessfully to get Sternberg to meet them.

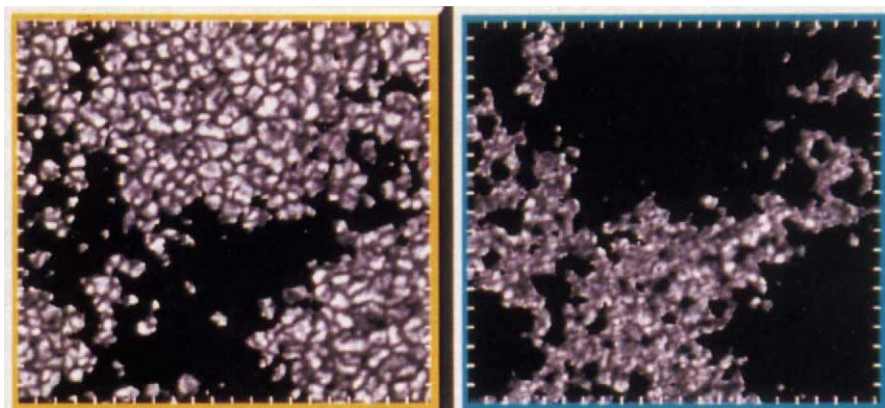
A congressional panel is looking into these legal manoeuvres as part of its continuing investigation into the L-tryptophan case. One aspect of the inquiry, by the human resources subcommittee of the House of Representatives Committee on Government Operations, is focused on whether Showa Denko's actions have hindered government investigations into the outbreak.

C.A.

You've read the paper, now see the video

These images of the surface of the Sun are stills from a short videotape that will be sent out with the 10 July issue of the *Astrophysical Journal*. All regular subscribers to the journal, the flagship publication of the American Astronomical Society, will receive a videotape, and subscribers outside the United States will be sent tapes in a format appropriate for their country.

On the tape are five segments related to papers in the journal. The solar images shown here form part of a movie that illustrates, in a series of high-resolution snapshots taken at roughly one-minute intervals, convective motion at the Sun's surface. The two images are complementary pictures of the same 30,000-km-square area showing regions of low and high surface magnetic field respectively. Although these still frames indicate that magnetic-field strength affects the convective cell structure, the movie as a whole, according to Alan Title of the Lockheed Palo Alto Research Laboratories in California, shows very plainly that the regions evolve differently over time: the less magnetized regions have a more pronounced



Areas of low (left) and high (right) magnetism on the Sun's surface.

cell structure, which changes on a faster timescale.

The use of movies in astronomy has been pioneered by theoretical astrophysicists performing computer simulations of, for instance, the formation and clustering of galaxies or the fate of matter spiralling into a black hole. Videotapes of such simulations have become a common feature of the annual meetings of the American Astronomical Society, and these movie presentations are much more than a gimmick. Both Title and Adrian Melott, a cosmologist from the University of Kansas whose simulation of galaxy formation is also in the current

videotape, say that the movies reveal features and processes not apparent in a series of stills.

The increasing use of videotape attracted the attention of Helmut Abt, editor of the *Astrophysical Journal*, who last year began sounding out a number of potential 'authors' of videotape material. Further video editions of the journal are planned as and when needed, the salient question being whether the tapes are essential to a paper or merely ornamental. Authors must pay a 'page charge' of \$1,000 for each videotape segment used.

David Lindley

Genentech's stance on biodiversity riles staff

San Francisco. A group of scientists from Genentech Inc. have publicly criticized their company's opposition to a biodiversity treaty adopted at last month's Earth Summit in Rio de Janeiro.

The US biotechnology and drug industries have argued that parts of the biodiversity treaty protecting genetic resources in rain forests and other natural habitats also threaten US leadership in biotechnology. In a letter of 9 June to the US president, George Bush, G. Kirk Raab, president and chief executive of Genentech, praised him for refusing to sign the document because of concerns about intellectual property rights and regulatory constraints.

The letter angered Anthony J. Pelletier and other scientists, who learned about their company's position from local newspapers. "Many, many people from all levels of the company were upset by the way this was handled", says Pelletier.

On 26 June, Pelletier wrote to Bush, calling his position "short-sighted" and saying that Raab's views did not represent those of all the staff and scientists at Genentech, nor the industry as a whole. "We see the loss of the rain forest as the loss of potentially great discoveries, and the concomitant loss of profits", Pelletier wrote. "Many of us in biotechnology see protect-

ing the diversity of life not merely as noble, but as good business as well."

Pelletier broadcast his counter-position within Genentech by electronic mail and word of mouth, garnering support from 70 scientists and staff members. He says that more employees would have endorsed his letter if his laboratory were easier to find and he had not been asked to stop using electronic mail after reaching only about one-fifth of the company's staff of 2,000.

Some of those who signed the letter — from secretaries and technicians to scientific directors — felt betrayed by Raab's failure to consult them. The South San Francisco company, a pioneer in the young industry, has long fostered a spirit of internal democracy and social-mindedness through such programmes as its Friday afternoon parties and an on-site childcare centre.

Raab said he acted independently — after consulting Genentech patent and law experts — because he felt strongly that Bush deserved commendation for making a risky and unpopular decision. "I didn't see it as a scientific issue," he said.

The letter of dissent has generated widespread discussion in the company. Many senior scientists have refused to take sides, saying they had no opinion "except the logical one — 'let's save the planet'". As one

scientist commented, "we're all too busy trying to figure out science, which is a lot harder than figuring out treaties".

The incident has prompted Genentech to go out of its way to clarify its support for both the concept of biodiversity and the right to individual expression. Company officials let local media know that Raab had written a letter on 2 July in support of a bill before the US Congress that would catalogue natural species and monitor their activity. Raab said late last month that he agreed with the intent of the biodiversity treaty and even with the "vast majority" of Pelletier's points. The convention, which was signed by 172 nations, gives them the right to control access to their own genetic resources and would ensure some sort of mutually agreed payment in return for profitable goods and patents created from those resources.

In an interview last week, Raab said he believes passionately in preserving biodiversity. "But, at the same time, I don't believe mixing in industrial property rights is the least bit appropriate. If you dig up a little piece of dirt in Naples ... or pick a flower in Ecuador, I don't think there is necessarily a requirement that the country has some predetermined economic rights", he says.

Sally Lehrman

Delays, confusion over rules hinder EC research projects

Munich. Bureaucratic obstacles are slowing progress in three research programmes sponsored by the European Commission (EC), frustrating applicants faced with unclear instructions and changing rules and angering those who succeed by making them wait for their money.

Most of the 15 programmes being funded under the commission's current research plan, which allocates ECU5.7 billion (US\$6.8 billion) over five years, are running smoothly. The plan, known as Framework 3, is intended to encourage cooperative efforts by requiring proposals from laboratories in more than one country. But there are problems with three programmes: the ECU131-million Biomed programme, which supports biomedical and health research; the ECU1.3-billion ESPRIT programme, which supports information technology; and the ECU488-million Human Mobility programme, which offers researchers the chance to move around Europe to gain expertise in techniques not available in their own regions.

The Biomed programme has been frustrated by feuds between the commission and its advisory committee over how applica-

tions will be peer-reviewed. The normal procedure is a two-tier process, with an initial screening to weed out the best applications followed by a more comprehensive review to decide funding.

In the case of Biomed, some 1,900 proposals are competing for only 100 grants. The programme's advisory committee, mostly heads of research councils or departments of health from each of the 18 participating countries, decided to invite 300 to submit full applications. But Filippo Pandolfi, the EC commissioner for research, changed the rules.

Bowing to pressure from groups that felt they had not been properly represented, Pandolfi asked the commission to give assessment scores to all applicants and invite them to proceed to the next step. This angered the committee, which unanimously refuses to back down from its earlier decision. It argues that a full review of low-priority proposals would waste the time of both scientists and reviewers.

More than 700 full applications have been received, but none has yet been put out for peer review. The reviewers were chosen for discipline-orientated assessment groups

last November, but they have not yet been called.

By contrast, proposed research under the ESPRIT programme has been peer-reviewed, but few of the successful applicants have received their money. The delay stems mostly from the commission's decision to go by the book after an unrelated decision that had not been ratified by all the commissioners was successfully challenged in January.

The administrators of ESPRIT, which received more than a thousand applications by its closing date last October, are now making sure that each grant decision is approved by each commissioner. So far, of the 300 successful applicants, 65 have had their grants fully approved and 140 more are expected to go out shortly. Two more batches will be completed by the end of the year. Officials at the commission say that they have also been slowed by the unexpectedly large number of applicants and the legal changes needed to administer research and development funding according to new rules written into the Maastricht Treaty, which prepares member countries for European unity.

The Human Mobility programme is intended as a way to distribute expertise throughout Europe; in practice, this means scientists from the poorer countries of southern Europe moving north. The start of the programme was delayed for several months, and when it was finally launched early last month applicants were given a deadline of only two to three weeks. In addition, the

China relaxes rules for visiting citizens

Beijing. China has adopted new rules that will make it much easier for Chinese nationals living and working in other countries to return for a visit. The changes all but abolish a system introduced shortly after the massacre at Tiananmen Square in June 1989 that severely restricted Chinese visitors wishing to return to positions outside China.

The decree, which took effect on 1 July, stipulates that anyone with a valid passport and an entrance visa to another country will be allowed to leave China freely. Mao Feng-Ping, director of the exit and entrance administration bureau in the Ministry of Public Security, says that the new rules apply "no matter how much time passes between the trip to China and the flight home and even if the country is different from where the visitor began his journey". The visitor need not obtain a visa inside China, nor need the passport be issued by mainland authorities. Under the former system, which required an 'exit registration card', such travel was nearly impossible.

Chinese scientists have long advocated the dropping of such restrictions because of the harm done to the country's research institutions. Talented young scientists who do not think they will be allowed to leave

China will choose to stay abroad. Similarly, the policy limits attendance at scientific conferences. For example, officials from the American Physical Society are concerned about attendance at an international conference on semiconductors in Beijing because of a lack of assurances from their hosts that all participants would be allowed to leave the country at the end of the conference, due to begin on 10 August.

The new rules will make life easier for Chinese researchers and students living outside China. The possibility of being retained during a visit to see family or attend a conference has exacted a high price from researchers who choose to live and work in another country. It was also a deterrent for Western scientists thinking about hiring Chinese nationals.

"When I heard that he wanted to go back to China for a visit before starting work here, I thought that I might never see him again", says Ellen Williams, a physicist at the University of Maryland, about Yunong Yang, a postdoctoral researcher who joined her laboratory a few months ago. "None of my Chinese students has ever had a problem, but you always worry."

In fact, Yang was forced to cancel his

travel plans after US officials told him that they needed six months to renew his passport. "They told me that they have to write to China and make sure everything is in order", he says, "and that sometimes they wait a long time before they get any answer. I couldn't wait that long." But Yang sent his wife and children, and they returned last month without incident.

Researchers within China are also happy with their newly granted freedom. For Li Ru-Jing, who completed his graduate training in Britain before returning to Beijing, the new rules may end an 18-month hiatus in his scientific career. "Now the ball is in my court", he says. "The test is whether I can find a foreign institute who wishes to take me in."

The change is part of China's campaign to open itself to the outside. The theme was sounded by party leader Deng Xiao-Ping during a tour of South China earlier this year and echoed by deputies at the National People's Congress in April. It may also reflect the country's need for hard currency from the West, along with a realization that excluding some of its most productive citizens is a poor way to stimulate economic growth.

Yau Qin Li

application forms from Brussels were so badly drawn up that nearly everyone requested further clarification.

The delay is due to confusion over how to incorporate political as well as scientific criteria into its assessment, with priority given to applicants from poorer countries. But EC officials cannot extend the deadline for applications because the ECU109 million available for 1992 will be lost if it is not spent by the end of the year. Late applications will be considered in 1993.

Although these problems provide a disincentive to many scientists, the lure of EC funds is irresistible. Young research fellows are offered salaries at nearly twice the going rate; at ECU45,000, they compare with those paid to department heads. Norman Bowery, professor of pharmacology at the University of London's School of Pharmacy, says that the short deadlines and ambiguous application forms put a lot of pressure on researchers, "but we can't not do it. There is so much money to be had."

In addition, researchers from poorer countries welcome the opportunity to compete for grants from outside their own governments. Rebecca Matsas, senior researcher in neurobiology at the Hellenic Research Institute in Athens, believes that the process is useful even if no grant is made. "It's hectic, but the interactions give you a chance to get to know a lot of people", she says. "It's good for future collaborations."

Alison Abbott

Scientific panel continues work on whaling ban

London. The moratorium on commercial whaling remains in place following last week's meeting of the International Whaling Commission (IWC) in Glasgow. Although a Revised Management Procedure that the commission's scientific committee had been working on for the past seven years was accepted (see *Nature* 357, 350; 1992), it must clear other administrative hurdles before it takes effect.

During the next 12 months, the scientific committee must develop minimum data standards, guidelines for conducting surveys and analysing the results and relevant computer programs. All this must be written into language that can withstand legal scrutiny.

Although a French plan to create an Antarctic sanctuary was withdrawn before the meeting opened, the commission decided to review the proposal for its next meeting, a year from now, in Japan. A working group of the scientific committee will hold meetings with such groups as the Commission for the Conservation of Antarctic Marine Living Resources and the Scientific Committee for Antarctic Research.

Ian Mundell

US oceanography lab sinks under weight of politics

Even congressional pork can go off. On 30 September, the US Navy will close its Institute for Naval Oceanography (INO) at Bay St Louis, Mississippi, thus dismantling part of the monument to himself that John Stennis built over a 42-year career in the US Senate. The collapse of the laboratory may be a warning that those who rely on powerful congressmen for launching new projects may find themselves stranded once those backers are gone or lose interest.

INO, the youngest of several naval projects at what is known as the Stennis Space Center, results from a deal struck in 1985 between Stennis, then a senior member of the Senate Armed Forces Committee and chairman of the Appropriations Committee, and John Lehman, then Secretary of the Navy. Stennis, angered by Lehman's decision to base in San Francisco a battleship he had coveted for Mississippi, was placated with the INO, planned as a bridge between the Navy and the academic research community. Among other things, INO would have used supercomputers and remote sensing data to forecast oceanographic conditions, much as meteorology has made weather forecasting possible.

Lehman's decision to locate INO in Mississippi contradicted the scientific advice and also the wishes of those nominated to work there, who favoured the Navy's research complex at Monterey, California. But Christopher Mooers, INO's first director, began on an upbeat note, with a staff of 30 and an annual budget of \$4 million and the ambition to attract others "with a pioneering spirit" (see *Nature* 324, 6; 1986).

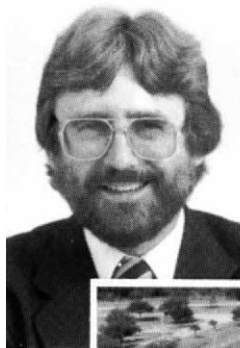
But it has since been downhill all the way. The budget has steadily declined to \$3.2 million, about two-thirds the value of that in 1986. The Cray Y-MP on which INO had set its heart went to the Oceanographer of the Navy instead (but INO researchers are allowed to use it). Mooers was forced to resign in 1989 after a management dispute, and since then there have been three directors. Mooers says that most of the original staff have left, like many of their successors. "Retention has turned out to be even more

difficult than recruitment", he points out.

Where does the blame lie? Most people give Stennis the lion's share, both for snatching INO from its preferred site in California and for his almost total lack of interest afterwards. "It was the barrel without the pork", says Mooers.

INO, created as a civilian centre to encourage cooperation between the Navy and universities, has been managed from the outset by the University Consortium for Atmospheric Research (UCAR), a group of 59 universities with interests in atmospheric research and oceanography. But the pairing proved to be a poor fit. UCAR's only members in southern states are from Florida and Texas. INO also suffered from the successive reorganizations of naval research, which have finally put it under the wing of the Naval Research Laboratory, which has its own university collaborations and has not

Christopher Mooers, former INO director (left); aerial view of institute (below).



been looking for UCAR's help.

Wayne Shiver, assistant to UCAR's president, says it has been clear from the start that the Navy would never provide the funds needed to let its newest and weakest research institution attain critical mass. In addition to insufficient funding, INO also suffered from the absence of a meaningful strategic plan and a vision of its place in the Navy's world. UCAR threw in the towel earlier this year, notifying the Navy that it wished to end its responsibility for INO. That was the death knell for the laboratory.

It remains to dismember INO, keeping the pieces worth saving. The ocean modelling programme will probably go to the University of Southern Mississippi and the Experimental Center for Mesoscale Ocean Prediction to Mississippi State University. The Navy will guarantee funds for two years, but then the programmes will have to compete for further grants. Most of the 31 people still at INO, including about 20 scientists, will be kept in one capacity or another. "We're trying to make the best of a bad situation", says Shiver.

Christopher Anderson

Head of US drug company takes Serbian post

Washington. If combativeness and an ability to stir controversy are what are needed to end a bloody civil war, then the Yugoslav government has found its man. Milan Panic, the founder, president and chief executive officer of ICN Pharmaceuticals Inc., in Costa Mesa, California, last week became prime minister of the remnants of Yugoslavia, which now includes the republics of Serbia and Montenegro.

Panic, a 62-year-old Serbian who immigrated to the United States in 1956, is no stranger to conflict. Over the past two decades, he has been embroiled in numerous confrontations with scientists, stockholders and the US government. ICN has been investigated by the Food and Drug Administration (FDA), a US House of Representatives oversight committee and twice by the Securities and Exchange Commission (SEC). More recently, Panic was widely condemned for promoting ribavirin, a drug developed by his company as an AIDS therapy; the FDA refused in 1987 to approve the drug for that use.

Those who know him say that it is Panic's personality rather than any real wrongdoing that has caused him problems. But even they agree that his aggressiveness and fiery temper may prove less than ideal for

his new position, in which he will be working with the Yugoslav president, Slobodan Milosevic, to form a new government.

Panic, who by his own account arrived in the United States with a family and only \$20, has spent the past 32 years building ICN into a multi-million-dollar corporation. Along the way, his tactics have made him the target of several government probes, the first in the 1970s, when ICN was accused of releasing inflated earnings potentials. The company settled with the SEC in 1977, not admitting to guilt but committing itself to new auditing policies.

The company gained even more notoriety in 1987 when it held a press conference to announce that ribavirin, which was developed by an ICN co-founder, had slowed the development of AIDS in men who had tested positive for the human immunodeficiency virus. The share price of the company's stock shot up, even though the FDA had approved the drug only for fighting respiratory infections in infants. Despite Panic's championing of ribavirin, the FDA did not approve the drug's use as an AIDS therapy, and some scientists view ICN's clinical trials with scepticism.

After all the publicity, the SEC began to investigate the company's promotional ac-

tivities, and shareholders filed a class-action suit accusing ICN of improperly publicizing ribavirin to raise the price of its shares. Last year, ICN settled with the SEC, agreeing to abide by the law but neither denying nor admitting to wrongdoing. ICN also settled with the shareholders, paying \$400,000 in penalties and \$200,000 in expenses.

Some observers familiar with the company say that Panic was guilty more of indiscretion than of falsely promoting ribavirin. They describe him as being fervently enthusiastic about his company and unconcerned about upsetting FDA officials. He is said to run his own company with a firm hand. "Here's an obviously enormously creative guy whose inability to negotiate and work with people who want different things and have different styles is his shortcoming," says Jonathan Kwitney, the author of a forthcoming book on the ribavirin controversy.

Panic has long had a keen interest in politics, hosting lavish fund-raising events and donating generously to Democratic politicians. Now that he has plunged into politics, it remains to be seen whether the traits that made him a controversial figure in the United States will help him to make peace in the Balkan states.

Traci Watson

New university funding council to reward potential

London. The new body funding higher education in England plans to broaden the way it assesses the quality and quantity of university research by taking into account the potential of a department and the size of its active research staff. This is a departure from the system now used by the Universities Funding Council (UFC), which decides quality almost exclusively on a review of existing work and relies on student numbers to measure research volume.

The new methods are discussed in a document released this week* by the UFC and the Polytechnics and Colleges Funding Council (PCFC), the two bodies that will combine in April 1993 to form the Higher Education Funding Council for England (HEFCE). The UFC this year will distribute £548 million (US\$850 million) and the PCFC £7.6 million. The councils will reach a decision in the autumn.

The new method, which will be applied to the 1993-94 funding round, will follow the current UFC model of using a formula that includes measurements of research quality and volume of activity. As such, it will make considerable use of an exercise now under way to assess research, and is expected to concentrate the research funding

further. For example, 25 institutions in England now receive 75 per cent of funds allocated for academic research.

The idea that departments with research potential should receive some sort of developmental funding reflects a concern on the part of the funding councils that a narrowly applied formula "would run the risk of ossification, and make it difficult for promising departments to develop their capacity". Although the amount of money involved is likely to be small, the move is being welcomed nonetheless.

Just how the research potential will be assessed is more problematic. A system that would judge a research plan submitted by a department would require the new council to have expertise it does not possess. The funding councils would like to use the existing approach that relies on a formula. A possible compromise might be to consider those departments that flunk the current assessment exercise but present a respectable research strategy.

The councils want to exclude departments with a high volume of research and whose parent institutions are otherwise well funded and in a position to help them. Departments that could be expected to get

funds from elsewhere are also likely to be excluded.

Under the UFC method, research volume is measured by counting the number of students in a department. Although the councils recognize that these students contribute to the research activity, they consider that, like other students, they are essentially working towards a degree. However, the contribution of research students will be taken into account in the assessment under way. The new approach will also be limited to active researchers.

The new council will retain the UFC's practice of measuring the amount of contract work done by a department. Although departments are expected to recover all the costs of contract research, there are some situations, such as research sponsored by the British government or the European Communities, where the work is fundamental and may need additional support.

Ian Mundell

* *The Funding of Research*, by the Higher Education Funding Council for England (available from the External Relations Unit, the Higher Education Funding Council for England, Northavon House, Coldharbour Lane, Bristol BS16 1QD).

Optical tests of Galileo's lenses

Vincenzo Greco, Giuseppe Molesini and Franco Quercioli

The science museum in Florence has two telescopes and a single lens attributed to Galileo. Tests conducted with modern interferometric equipment show that Galileo was able to obtain nearly perfect optical quality.

ACCORDING to biographies, Galileo is supposed to have made several telescopes, purchasing some lenses and polishing others himself, yet only the optical apparatus now in Florence, collected and handed down by the Medici family¹, appears to have survived. The authenticity of the telescope tubes seems certain, but some doubts remain about the lenses, apart from the single lens, which was used in the discovery of the Medici stars.

The Medici collection was examined in 1923^{2,3}, but we have now taken the telescopes apart and tested them with state-of-the-art optical equipment. Analysis of the optical quality of Galileo's lenses is of interest for understanding both the development of optical technology and the observational capabilities of early astronomers. Before Galileo's time, observations were made with the naked eye, for which the resolution is about 1 arc minute. Lenses were used only as eyeglasses or as magnifying lenses, for which applications the poor glass purity and optical figure then feasible were nevertheless adequate.

We have examined Galileo's optics with modern methods. We call the telescopes I (paper coated, longer tube) and II (leather coated, shorter tube). Their optical configuration is based on a positive objective and a negative eyepiece

and objective II (*c* in the figure) are really plane to a fraction of a wave, although flatness of these surfaces is not required in terms of image quality. (Polishing a surface to such flatness is not trivial even for today's crown

OPTICAL DATA OF GALILEO'S LENSES

	Front radius	Back radius	Central thickness	Full diameter	Aperture diameter	Focal length
Objective I	2,700	950	2.5	51	26	1,330
Eyepiece I	plane	48.5(*)	3.0	26	11	-94.0
Objective II	535(*)	Plane	2.0	37	16	980
Eyepiece II	51.5(*)	51.5(*)	1.8	22	16	-47.5
Single lens	940	12,000	4.0	58	38	1,710

Asterisks, data from ref. 3. Dimensions are in mm.

with a common focus. The table summarizes the measurements of the geometry and the first-order optical parameters. The focal lengths are measured in the centre of the visible spectrum (550 nm). From the lens geometries and focal lengths, we calculate that the refractive index of the glasses is 1.51–1.55. The relative apertures of the objectives are $f/51$ for telescope I, $f/61$ for telescope II and $f/45$ for the single lens. The magnification of the telescopes is 14 for I and 21 for II.

We used a 633-nm digital phase-shift Fizeau interferometer to study the regularity of the optical surfaces and the wavefront distortion in transmission. Typical fringe patterns are shown in the figure. As far as regularity is concerned, the quality of the objective lenses is far better than the quality of the eyepieces. But because the used diameter per field angle at the eyepiece is much smaller than the clear aperture of the objective, the effect of the lower quality of the eyepieces is negligible. It is surprising that the plane surfaces of eyepiece I

glasses.)

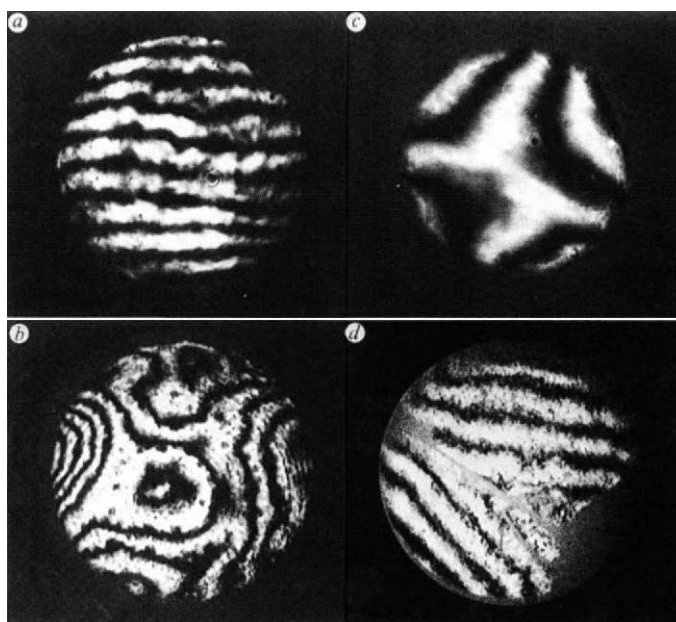
A further observation comes from the appearance of the concave surface of eyepiece I (*b*). In addition to the interference fringes, a pattern of ring shadows appears, as if the surface had traces of a turning process. The wavefront distortion of objective I is very small (*a*). The best quality belongs to the single lens, which can be considered as nearly diffraction-limited (*d*). According to the Rayleigh criterion, its resolution at 633 nm is of the order of 3 arcseconds. Of course, the optical performance of the telescopes is degraded for several reasons, mainly chromatic aberration. Computer simulations taking dispersion into account lead to estimates of only 10–20 arcseconds resolution over the visible spectrum.

Altogether, our tests of the lenses (made 350 years after Galileo's death) show that they are polished to a good spherical shape, and the presence of proper apertures on the objectives also shows Galileo's awareness of the need to tune the optical performance. As a result, although affected by intrinsic chromatic aberration, at single wavelength the telescopes are nearly diffraction-limited, that is, optically perfect. □

Vincenzo Greco, Giuseppe Molesini and Franco Quercioli are at the Istituto Nazionale di Ottica, Largo E. Fermi 6, Firenze 50125, Italy.

ACKNOWLEDGEMENTS. We thank P. Galluzzi and M. Miniati for discussions, and the Istituto e Museo di Storia della Scienza for permission to publish the material in this paper.

1. Bonelli, M. L. in *Proc. 1964 Symp. Galileo* 125–127 (Marzocco, Florence, 1967).
2. Abetti, G. *L'Universo*, **4**, 685–692 (1923).
3. Ronchi, V. *L'Universo*, **4**, 791–804 (1923).



Fizeau fringe patterns of the optical elements of Galileo's telescopes at 633 nm. The fringe maps show the deviation of the wavefront from a sphere or a plane. For reference, a diffraction-limited wavefront produces no fringes, or straight and equally spaced fringes if some tilt is added. *a*, Double-pass interferogram of objective I, folded with a reference mirror. Deviations from straight fringes are of the order of half a pitch, meaning a departure from the ideal wavefront of the order of a quarter of a wavelength. *b*, Reflection interferogram of the concave surface of eyepiece I. Fringes are highly irregular. *c*, Reflection interferogram of the plane surface of objective II. The quadrant fringes show astigmatism. *d*, Double-pass interferogram of the single lens.

London Zoo crippled

SIR — Your leading article (*Nature* 357, 613; 1992) repeats familiar but unsubstantiated charges against the council of the Zoological Society of London. As Lord Zuckerman points out in the same issue (357, 621; 1992), cash flow problems began in the 1960s and continued through to the 1980s. Acknowledging London Zoo's status as a national institution, government gave occasional grants to restore operating losses, while other national zoos received large annual subsidies. The zoo management has never been perfect but few if any capital zoos can match London's record of raising 80 per cent of its costs from visitor revenues.

You ignore Zuckerman's clear statement that the 1988 settlement of £10 million from the government was conditional upon acceptance of a new management with, as you say, "an entrepreneurial cast of mind". The council had little part in this — even the chief executive was chosen by the government.

As for pricing ourselves out of the market, where else can you get a full day out for £6, the price of a cinema seat? The *Report on London Zoo* published by the Select Committee on the Environment in June 1991 gives comparative entrance prices for other visitor attractions and London Zoo is shown to be highly competitive. Few zoos today make a profit, and it is disingenuous to make unfavourable comparisons with provincial zoos whose situation is utterly different.

Like many commentators, you seem to equate London Zoo with the Zoological Society. Much as we all wish the zoo to survive, it is not indispensable for the work of the learned society, the library, the scientific meetings, publications, the research laboratories and of course Whipsnade. It is much to the credit of the government that it appreciated the distinction in 1988 and guaranteed an annual grant of £1.3 million (indexed) to support the research of the Institute of Zoology four years before biodiversity became a political byword.

Barry Cross

(Secretary)

Zoological Society of London,
Regent's Park, London NW1 4RY, UK

No connection

SIR — There is much talk these days of irresponsible and misleading reporting in the media and I had not imagined that I should be writing to level such a charge

against *Nature*. But your leading article "Leeds Disunited" (357, 614; 1992) leaves me no option. It refers to an allegation about scientific misconduct reported to me last September that led ultimately to an investigation by a trio of eminent scientists, two from this university and one from an internationally renowned medical school in London. The panel followed guidelines modelled on those recommended by the Royal College of Physicians and it reported in March. The essence of the Royal College guidelines is that any such inquiry should be confidential.

In the event, the panel reported that the balance of evidence supported the view that there had been fraud (albeit of a minor nature) by a junior scientist and I wish to say only that appropriate action has been taken by the university. As far as we are concerned, the matter ends there.

Any attempt to relate this incident by inference to the entirely separate situation at Leeds General Infirmary betrays a *post hoc ergo propter hoc* style of reasoning that I would have thought inimical to the scientific standards espoused by your journal. The remarks in the last paragraph of your leading article are likely to work against the scientists whose integrity you appear to wish to protect.

J. J. Walsh

(Registrar)

The University, Leeds LS2 9JT, UK

Nonpersons

SIR — If we are led to believe (*Nature* 357, 425; 1992) that an eight-cell blastomere cannot be considered a person, then as a Christian I believe this to be totally incompatible with the biblical text concerning the Incarnation. To suggest that God become man at some arbitrary future date would be ridiculous.

Anthony Gannon

Division of Quantum Metrology,
National Physical Laboratory,
Teddington, Middlesex, UK

Journalists in labs

SIR — Jonathan Stamford (*Nature* 357, 10; 1992) comments on the need to find means of placing journalists in suitable laboratories for a period of weeks, so that they "might then appreciate not only how scientists think and interact but the nature of the research process".

Your readers will be interested to learn that a programme designed to do exactly what Stamford suggests will be initiated in April 1993 at the Max Planck Institutes of Biochemistry and Psychiatry in Munich, with the endorsement of the

European Communities, the Committee on the Public Understanding of Science, the Max Planck Society and similar high-level research agencies in the major European countries. The programme is called EICOS (European Initiative for Communicators of Science) and will be open to established science reporters from all the countries of Europe and all branches of the media (newspapers, magazines, radio, television and freelance), as well as to editors and broadcast producers.

Byron H. Waksman

(Program Director, EICOS)

Science Writing Fellowships Program,
Marine Biological Laboratory,
Woods Hole, Massachusetts 02543, USA

Einstein's beliefs

SIR — Many people asked Albert Einstein about his religious views and I, when a US Navy ensign in 1945, was one of them. On 2 July 1945, in response to a letter from me quoting an acquaintance who claimed that a Jesuit priest had convinced Einstein that a supreme intellect governs the Universe, Einstein denied ever having spoken to a Jesuit priest and said: "From the viewpoint of a Jesuit priest I am, of course, and have always been an atheist. . . ."

Five years later, on 25 September 1949, I wrote again: ". . . [In your letter,] You say that 'From the viewpoint of a Jesuit priest I am, and have always been, an atheist'. Some people might interpret that to mean that to a Jesuit priest, anyone not a Roman Catholic is an atheist, and that you are in fact an orthodox Jew, or a Deist, or something else. Did you mean to leave room for such an interpretation, or are you from the viewpoint of the dictionary an atheist; i.e. 'one who disbelieves in the existence of a God, or Supreme Being?' . . ."

Einstein's response, dated 28 September 1949, says: ". . . I have repeatedly said that in my opinion the idea of a personal God is a childlike one. You may call me an agnostic, but I do not share the crusading spirit of the professional atheist whose fervor is mostly due to a painful act of liberation from the fetters of religious indoctrination received in youth. I prefer an attitude of humility corresponding to the weakness of our intellectual understanding of nature and of our own being."

Guy H. Raner Jr

22244-2 James Alan Circle,
Chatsworth, California 91311-7137, USA
with the assistance of

Lawrence S. Lerner

Department of Physics and Astronomy,
California State University, Long Beach,
Long Beach, California 90840, USA

Long-range correlations within DNA

The idea that nucleotide bases in strands of DNA may be correlated over several thousands of base-positions has been established, but has not yet been explained.

ARE there long-range correlations between the nucleotides within a gene and, if so, what purpose can they serve and how, in any case, do they arise? These are the questions undoubtedly prompted in many minds by the publication in March this year, by C.-K. Peng of Boston University and his colleagues (*Nature* **356**, 168; 1992), of an article demonstrating long-range correlations extending over many thousands of base positions in several different DNA sequences. It was natural to expect that the noise level would soon be filled with the clash of opinions between sceptics and believers, no doubt equipped with at least one supposed cause of error, or with an explanation as the case might be. Instead, surprisingly, it is the silence that has been deafening.

So far, at least. But now Richard F. Voss from the IBM Research Center at Yorktown Heights, New York, has responded in *Physical Review Letters* with what may be a more explicit way of calculating the correlations (**68**, 3805; 22 June 1992). The results are inherently the same — there are indeed long-range correlations along the length of real-life stretches of DNA. The long-range correlation may even differ significantly between taxonomic groups. But nothing is yet said about their origin or function.

The expectation that there will be no long-range correlations is natural, and is the obvious baseline against which correlations in real DNA are measured. That there may be short-range correlations in DNA is another matter: crude expectation would suggest there might well be. Armchair molecular biologists will quickly suggest several plausible patterns: purines (adenine and guanine) may alternate with pyrimidines (thymine and cytosine), purines (or pyrimidines) may alternatively predispose to successors along the chain of their own kind, or these patterns may occur in different parts of a gene according to its function. But nobody would expect that correlations such as these hypotheses imply would extend over more than a few dozen bases, corresponding to nearest neighbour interaction between nucleotides along the length of a twisted double helix.

So how to calculate the chance that the illusion of long-range correlation will arise by pure chance? By the calculus of the random walk, as used in Einstein's theory of Brownian motion. If, for example, the issue is simply whether there is a correlation between the occurrence of a purine base at one point in a stretch of DNA and a purine (or a pyrimidine) base 10,000 base-positions further along, the first

step is to know what chance would throw up. One could imagine (as Peng *et al.* did) the structure of a particular gene plotted on a one-dimensional graph — one step to the right for a purine, one step to the left for a pyrimidine. Then the problem is simply that of Brownian motion in one dimension.

If chance is all that matters, the result is straightforward. The likelihood that there will be a purine here and also a purine there will decrease steadily (monotonically, as they say) with the distance between here and there. For a strictly random walk, the square root of the average of the square of the displacement of the point on the walking graph from its average position (the root mean square displacement) will be proportional to the distance to the power 0.5. Another way of putting that (after a little algebra) is that there will be a gaussian distribution of the position of a particle along a one-dimensional lattice which becomes more widely spread out as time passes (or as the distance between base-positions increases). Any other kind of behaviour implies that some nonrandom process is under way.

So how best to test that the long-range correlations between nucleotides in DNA differ meaningfully from these expectations? There are obvious pitfalls. To deal separately with the four bases that crop up in real DNA, it is essential to consider the process of random-walking in four dimensions; otherwise, correlations that are artefacts will crop up. And what happens when there are external constraints in the analysis of a particular stretch of DNA, as when it is richer in C and G than in A and T (or conversely)?

Voss's way has an intrinsic simplicity, guided by the binary character of the arithmetic of the computer. Represent the sequence of a gene by four separate sequences, each referring to one of the nucleotides that ordinarily crop up; the sequence contains a '1' if a site is occupied by that nucleotide concerned, otherwise a zero. Adding the four sequences together then gives a simple sequence of '1's. From these, it is then possible to calculate all possible correlations for a given DNA sequence, say the chance that there will be an 'A' at one position and a 'C' exactly fifty positions further along the strand, and to take the average over all possible placings of the first position. Readers of Voss's beguiling article should be warned that there is some fancy binary algebra embedded in his proof of the link between the elementary sequences and that from the real world.

Voss's biggest single achievement is

to have run the complete genome of cytomegalovirus (strain AD169) through his computer, but altogether he has analysed more than 50 million nucleotide position in the more than 25,000 DNA sequences distributed as GenBank Release 68. The outcome is quite remarkable. There are indeed long-range correlations, extending over thousands or more of base positions.

Voss hammers home the point that the character of the long-range correlations is that of what is called '1/f noise', which is shorthand for the notion that a noise-level may be a fractional power (not very different from unity) of 1/f, of the kind that arises in systems of a fractal character among others. But using the fractional power as an index (say β), it emerges that there are significant differences (ranging from 0.76 to 0.92) in the values of β describing the long-range correlation between the occurrence of the same bases.

In passing, the speculations of the armchair molecular biologists are confirmed. In cytomegalovirus, pyrimidines are anti-correlated at some sites while at others the opposite is the case. One intriguing feature of the analysis is that the triplet DNA code shows up as a bump in what would otherwise be the 'high-frequency' end of the spectrum where the frequency corresponds to a spacing of three units. It is also interesting that Voss has used as a random 'control' a sequence consisting of the first 1.13 billion decimal digits of the number π .

The evolutionary comparisons of long-range correlations are the most intriguing. Both phage and bacteria have β roughly equal to 1.0, but the spectral density does not vary linearly (on a long-log plot) as for other taxonomic groups. Vertebrate sequences yield a similar slope, but a more regular one. The lowest values of β are for primates and organelles (such as mitochondria). In all groups, there is a peak of short-range correlation corresponding to a three-nucleotide displacement, but in vertebrates, invertebrates and primates (but not rodents or mammals in general) there is also a peak corresponding to a displacement of nine units.

Voss is modest about his work, describing it as "only a beginning". He can say that again. No doubt others are already preparing to follow where he has begun. But still, there is no explanation of these long-range correlations. Are evolutionary processes at work? Or may the long-range correlations arise because genes in higher organisms have been derived by patching together genes from simpler entities? **John Maddox**

Cold dark matter under a cloud

James Binney

APRIL was a bad month for the standard cold-dark-matter (CDM) cosmology. Inspired by highly speculative ideas in particle physics, the CDM model has had significant successes in modelling the evolution of structure in the Universe¹. But under attack by studies of the large-scale clustering of galaxies², its health has recently been giving cause for concern, and in April it suffered further set-backs — first from the magnitude of the primordial fluctuations in the cosmic microwave background detected by the COBE satellite³; and second, from Tóth and Ostriker⁴, who argue that the existence of thin galactic disks like that of the Milky Way is incompatible with the existence of cold dark matter.

COBE's measurement of surprisingly large primordial fluctuations on the scale of 10° and above attacks CDM where it has long been weakest — on the largest scales. So it just adds weight to the long-standing criticism that standard CDM theory is unable to account for the observed degree of large-scale clustering in the Universe. Moreover, the interpretation of the COBE measurement is far from straightforward on account of uncertainties in the contributions to the measured fluctuations from dust, bremsstrahlung and synchrotron emission from within the Milky Way. The argument of Tóth and Ostriker is interesting because it attacks CDM where it has traditionally been strongest — on small scales.

The CDM model assumes that the Universe has the critical mean density ρ_c associated with the flat Einstein-de Sitter cosmological model; in this model the Universe expands for ever, but the expansion rate tends to zero rather than stabilizing at a finite value as in cosmological models with lower mean densities. Because ρ_c exceeds the density one would get by spreading out the masses of the visible galaxies uniformly, the CDM theory holds that more than 90 per cent of all mass lies between the galaxies. In the simplest picture, each galaxy is assigned its share of this matter and one imagines that every galaxy is enveloped in a massive halo a megaparsec or so in extent, in which the density falls with radius r roughly as $\rho \propto r^{-2}$. Measurements of the velocities of clouds of neutral hydrogen⁵ in orbit at radii of 30 kiloparsecs or so, and of X-ray emitting gas⁶ at similar distances from giant elliptical galaxies, provide direct evidence for the existence of such dark halos.

These halos cannot be static things, however. To see this, consider the fate of a small over-dense region in an otherwise perfectly homogeneous Einstein-de

Sitter universe. In the Einstein-de Sitter model, the recession velocity of any particle, A, from any other, B, is exactly equal to the escape speed of A from the sphere of material around B whose radius is equal to the distance of A from B. Thus as the Universe expands and the sphere around B becomes larger and less dense, the recession velocity of A from B tends gently to zero. But if the density in the neighbourhood of B exceeds ρ_c by some amount, no matter how small, A's recession velocity will vanish when the sphere has a finite size, and A will then fall back towards B. In essence the Einstein-de Sitter model is unstable, and any upward perturbation in the density around some point will cause the whole Universe eventually to crash down on that point. Galaxies are thought to form around maxima in the cosmic density field, so it looks as if the CDM model predicts that the whole Universe will eventually crash down on us here in the Milky Way.

Tóth and Ostriker's argument against CDM turns on how true this naïve expectation is. They start by pointing out that galactic disks are fragile objects which are easy to break and hard to mend without the damage showing. This situation arises because in a galactic disk like that of the Milky Way, stars are concentrated on less than 1 per cent of the orbits energetically open to them, namely onto nearly circular orbits that are inclined by less than a degree to the galactic plane. This concentration would be reduced by the gravitational field of any lump of matter which passed through the disk, because such a field would scatter stars more or less randomly and so, on average, move stars from more to less highly populated orbits. Macroscopically, the intruder would have the effect of making the disk thicker and 'hotter' in the sense of having a larger radial velocity dispersion.

In fact older stars near the Sun form a thicker and hotter disk than do the youngest stars, and this fact has long been interpreted as evidence that stars are gradually scattered from the highly circular and planar orbits on which they were born by irregularities in the Galaxy's gravitational field⁷. The precise nature of these fluctuations has, however, remained obscure. Giant molecular clouds and spiral arms must both make useful contributions, but it seems that neither contribution is adequate to explain the noncircular motions of the oldest disk stars. Scatterers that are not confined to the disk seem to be called for, and infalling lumps of dark matter

may be just what the doctor ordered. By Tóth and Ostriker's estimate, the distribution of stars in the solar neighbourhood could be explained if just 3–4 per cent of the Galaxy's mass interior to the solar orbit had been accreted as lumps since the disk formed about 5–10 billion years ago.

How does this 3–4 per cent compare with the expected accretion rate in the CDM picture? It is hard to say. The easy part of the calculation is estimating how rapidly mass falls in on the Galaxy — it does so extremely rapidly; the Galaxy's mass is thought to have roughly doubled in the past 5 billion years. What is very much harder is estimating how lumpy the infalling matter is, and how many lumps reach the solar neighbourhood before being torn apart by the Galaxy's tidal field.

In the CDM model galaxies form 'from the bottom up' through the merging of small proto-galaxies to form larger proto-galaxies, and then of these larger proto-galaxies to form still larger proto-galaxies and so on. So much of the expected infall will come in huge lumps such as the Magellanic Clouds, which the Galaxy appears to be tearing apart even now, and the Andromeda nebula, which is approaching us and looks set to tear us apart within the next 10 billion years. Other material will come in smaller lumps, perhaps ones that contain no stars. Fortunately for the disk, all objects will be greatly reduced in mass as they fall through the halo, and only if their central densities are higher than the mean density (about one atom per cubic centimetre) of the Galaxy interior to the solar circle will they reach the disk. Tóth and Ostriker assume that near their centres the density of lumps rises as r^{-2} , so meeting this condition, and conclude that CDM predicts that 90 per cent of all disks would have been disrupted within the past 5 Gyr. This conclusion is hard to square with the fact that the majority of luminous galaxies have old, cold disks.

Although the argument is interesting, it is far from conclusive because we cannot yet reliably calculate the characteristics of infalling dark matter. The two key characteristics are the mass-spectrum of the lumps, and the density profile of an individual lump. Very small lumps will orbit in the halo for a Hubble time (the age of the Universe) and not disturb the disk, while more massive

1. Frenk, C. S., White, S. D. M., Efstathiou, G. & Davis, M. *Astrophys. J.* **351**, 10–21 (1990).
2. Sanders, W. et al. *Nature* **349**, 32–38 (1991).
3. Silk, J. *Nature* **356**, 741–742 (1992).
4. Tóth, G. & Ostriker, J. P. *Astrophys. J.* **389**, 5–26 (1992).
5. van Albada, T. S., Bahcall, J. N., Begeman, K. & Sancisi, R. *Astrophys. J.* **295**, 305–313 (1985).
6. Fabian, A. C., Thomas, P. A., Fall, S. M. & White, R. E. *Mon. Not. R. astr. Soc.* **221**, 1049–1056 (1986).
7. Jenkins, A. & Binney, J. *Mon. Not. R. astr. Soc.* **245**, 305–317 (1990).

lumps will be dragged in to the radius of the disk by dynamical friction against the halo. Lumps that do not possess sharp density cusps (peaks) will either be tidally torn apart in the halo or are so reduced in mass that they will orbit in the halo for more than a Hubble time.

In the CDM picture the central density profiles of lumps of dark matter might be expected to follow some power law in radius. The slope of this power law is hard to estimate at present and may differ from that assumed by Tóth and Ostriker. The existence of clusters of

galaxies argues against a steep profile, as it would prevent individual galaxies surviving the virialization of a cluster, leading to the formation of one vast, amorphous galaxy rather than the clusters we observe. All the evidence points to small halos being too fragile to survive infall into larger ones. It is not clear that this fragility is incompatible with CDM, but it does look a distinct possibility. □

James Binney is in the Department of Physics, University of Oxford, 1 Keble Road, Oxford OX1 3NP, UK.

MOUNTAIN BUILDING

Crust in mantle overdrive

Peter Molnar

THERE is little doubt that the engine that builds mountain ranges lies in the mantle, fuelled by density differences between cold and hot mantle material and driven by gravity. Most mountain ranges result from thickening of crust where blocks of crust converge towards one another. Through isostasy, the Earth's version of Archimedes' principle, the thickened crust is buoyed up to form a high terrain. What remains the object of much search and speculation is the mechanism by which the convection in the mantle drives the crustal blocks together. A simple and sensible view holds that the cold mantle lithosphere directly beneath the belt must also thicken during convergence, and that this thickened root, being heavier than the surrounding hotter mantle, sinks and draws the neighbouring blocks towards one another (*a* in the figure).

But on page 144 of this issue¹, Makeyeva, Vinnik and Roecker claim that the mantle follows a different agenda. They suggest that instead of converging towards the belt and then flowing downwards, much of the flow in the mantle directly beneath the belt is parallel to it (*b* in the figure), and that in some places mantle actually upwells beneath a region where crustal shortening and thickening continue.

The authors analysed seismograms recorded in the Tien Shan, a major intracontinental mountain range in central Asia. Unlike many large mountain belts (such as the Himalayas or Alps), formed where two continents have collided with one another after a tract of oceanic lithosphere was subducted, the Tien Shan is truly intracontinental. The last oceanic lithosphere disappeared more than 300 million years ago, and the whole region seems to have been eroded to a featureless plain before renewed mountain building began just 30 million years ago.

So the asymmetric processes that characterize subduction zones and collisional belts may not complicate the deep structure of the Tien Shan. Moreover, the range is linear, and if one were to search for an active mountain belt in which variations in structure, and presumably processes, along the belt would be negligible compared with the variations across it, one could not find a better example. Yet the linear topography and surface geology of the Tien Shan conceal large variations in mantle structure along the belt.

Building upon several years of investigation by Kazakh, Kyrgyz and Russian seismologists, Makeyeva *et al.* claim that the mantle structure beneath the part of the Tien Shan west of the border with China is not underlain by a cold lithospheric root at all, but rather by material with low seismic-wave velocities, and hence presumably relatively hot material. Ultimately, they argue for an upwelling of mantle material beneath this part of the range, but the crucial evidence in their argument is provided by an analysis of seismic anisotropy in the mantle beneath the region.

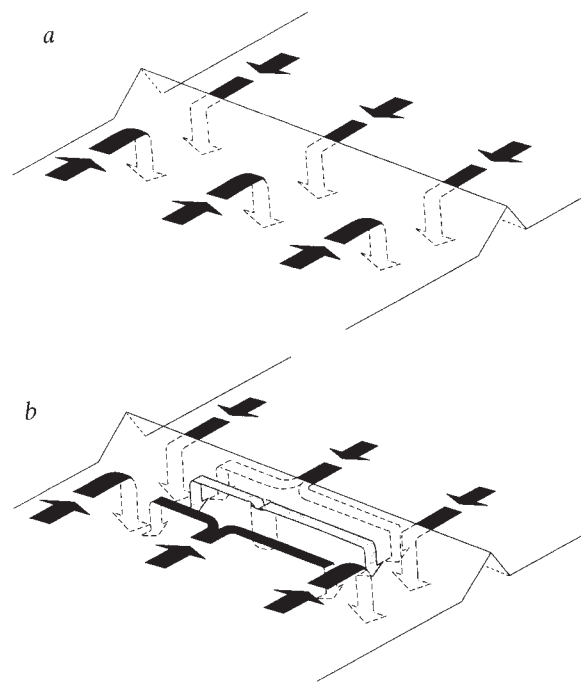
Anisotropy has, with a few notable exceptions, traditionally served one of two ignominious roles in seismology. Because tedious algebra is needed to describe wave propagation through anisotropic media, anisotropy for many seismologists has been a nuisance that inhibits simple analysis, interpretation and

intuitive understanding, and it commonly is ignored. For others, it has been a *deus ex machina*, introduced to rescue a set of otherwise inconsistent observations. But even under these circumstances, it often fails to provide any insights into the interesting processes within the Earth.

Makeyeva *et al.* exploit anisotropy to convert a seemingly exceptional mantle structure beneath the Tien Shan into a rule: one that bears on the significance of anisotropy as a tool, as much as it does on the processes of mountain building.

They show clearly that the faster orientation of the two essentially horizontal polarizations of S (secondary or shear) waves propagating up through the mantle beneath the Tien Shan lies roughly parallel to the belt. This parallelism is common in other active mountain belts. Yet, in the area of the unusually hot mantle material, this orientation deviates markedly from the easterly trend, again making the area exceptional. The casual reader might gain the impression that anisotropy merely provides another entry in a list of anomalies from this area.

As Makeyeva *et al.* state clearly, anisotropy can arise in a variety of situations, and interpreting it in terms of dynamic processes must contain an element of speculation. Nevertheless, citing a number of studies, they assume that the orientation of the faster axis, which



a, Simple pattern of mantle convection that might build intracontinental mountain belts, with linear downwelling beneath a cold lithospheric root. *b*, Circulation pattern indicated by the results of Makeyeva *et al.*, running parallel to the chain and with an upwelling plume beneath the chain.

need not lie parallel to the fastest of the three axes², lies parallel to the orientation of maximum shear. That assumption makes the rule that the relatively fast axes aligned parallel to mountain belts imply flow in the mantle parallel to them. It follows that the region of anomalous mantle becomes the locus of the upwelling limb of convection parallel to the belt, and that such convective flow characterizes active belts in general.

The application of anisotropy to questions of mantle structure and dynamics is not new in the Earth sciences, but the study by Makeyeva *et al.* illustrates how this wanton, outcast child of seismology can be tamed and put to good use. The inclusion of anisotropy in the toolbox of most seismologists, however, probably will not happen without the smashing of some sacred icons. For instance, there is already a suggestion that much of what has been interpreted as lateral variations of isotropic velocity structure, and hence of temperature in the mantle, might instead be due to lateral variations in anisotropy^{3,4}.

The variations in mantle structure along the Tien Shan are not exceptional. Even for the most celebrated inference of cold material sinking beneath a convergent mountain belt — the Western Transverse Ranges of California^{5,6} — the deep structure is poorly described as two-dimensional. Variations in mantle structure along the Alps are as large as those across it⁷, and the vast flat surface of the Tibetan Plateau conceals lateral variations in mantle structure as large as any beneath continental regions^{8,9}. Thus, the image of a lithospheric root mirroring the overlying topography of a mountain range may owe its persistence more to belief than to fact.

Makeyeva *et al.* touch only briefly on the process that may cause flow and sense of shear in the mantle to be perpendicular to that observed by geologists in the crust of all mountain belts. The sinking of a linear belt of cold material, the nascent lithospheric root, creates a convective instability, because the material is more dense than its surroundings. The growth of the instability may be faster for flow parallel to the chain than perpendicular to it.

A similar, three-dimensional instability has been observed in numerical experiments of flow at mid-ocean ridges, for convection enhanced by melt segregation and for sufficiently fast spreading¹⁰. Analysis of corresponding three-dimensional instabilities for lithospheric roots may reveal a similar process beneath mountain belts. Fortunately, these problems do not yet seem to have interested those numerical mod-

ellers more interested in simulating complicated situations than understanding simple ones. Thus, the possibility of gaining a simple understanding of the mantle dynamics beneath mountain belts remains open, instead of confused. □

Peter Molnar is currently at the Laboratoire de Géophysique Interne and Tectonophysique, BP 53X, Université Joseph Fourier, F-38041 Grenoble Cedex, France.

MOLECULAR BIOLOGY

Chlamydomonas surrenders

Karen P. VanWinkle-Swift

WHOLE-organism model systems are the bedrock of much of modern biology. Yeast and *Drosophila* yielded long ago to the attempts of molecular biologists to isolate, manipulate and reintroduce homologous (or even foreign) DNA sequences into their models — thus turning the classical genetic strengths of these systems into molecular strengths as well. That the unconditional surrender of the unicellular alga *Chlamydomonas* (sometimes irreverently referred to as 'the green yeast') is now close at hand became plain at a recent meeting*. The problems being addressed and the tacks being taken are as diverse as biology itself — characterizing microtubule-associated motors, deciphering photosensory transduction pathways, unraveling communication networks between nucleus and chloroplast, to name just a few. But evidence for the molecular surrender pervaded the conference.

An earlier reunion of *Chlamydomonas* researchers had included both the exciting report of high-frequency nuclear gene transformation¹ and the disappointing news that gene integration was nonhomologous (as opposed to the typically homologous integration of transforming organelle DNA²). The persistent cohabitation of introduced and resident nuclear gene copies (and the typical dominance of resident alleles) prohibits the molecular dissection of gene function by site-directed mutagenesis and gene replacement. But yesterday's bad news turned into the good news of the 1992 meeting: reasoning that nonhomologous integration should often occur within other genes of interest — and thereby prevent their function — several investigators are now using transformation as a form of mutagenesis simultaneously to disrupt and tag particular genes.

For example, motility mutants can be easily obtained when wild-type transformants, recovered after transformation of

a strain that does not use nitrate (*nit1*) with the cloned nitrate reductase (NR) gene³, are screened for swimming defects (L.-W. Tam, University of Minnesota). Similarly, photosynthesis mutants are readily recovered from an arginine auxotrophic strain transformed to arginine-independence by the cloned argininosuccinate lyase (ASL) gene⁴ (M. Goldschmidt-Clermont, University of Geneva). Support for the assumption that the new mutant phenotype is due to insertion of the transforming DNA comes from experiments, involving crosses to wild-type strains, in which the mutant phenotype (for instance a motility or photosynthesis defect) cosegregates with the selected transformed phenotype (the ability to grow on nitrate, or without exogenous arginine).

If the transforming event includes insertion and maintenance of vector sequences as well as the originally selected *Chlamydomonas* gene (which unfortunately is not always the case), DNA flanking the insertion (that is, the disrupted gene) can be cloned by plasmid rescue. This technique takes advantage of an antibiotic resistance marker which is present in the original vector and can be selected by transformation of *Escherichia coli*. The flanking DNA can then be used to select the full-length gene from genomic libraries. The approach has been carried to completion with the successful cloning of a nuclear gene involved in the unusual trans-splicing of the chloroplast *psaA* transcript⁵, the absence of which causes a photosynthesis-defective mutant phenotype (M. Goldschmidt-Clermont).

This method of gene cloning is ready for application to any of the many areas of *Chlamydomonas* research in which mutant phenotypes already define functions of interest — in other words, it is applicable to all of *Chlamydomonas* biology. One simply needs to be willing to reisolate the phenotype using transforming DNA as the 'mutagen', and for most of those in the business that pursuit is

1. Makeyeva, L. I., Vinnik, L. P. & Roecker, S. W. *Nature* **358**, 144–147 (1992).
2. Plomerová, J. *et al.* at Conf. Eur. seismol. Comm., Prague (1992).
3. Babuška, V., Plomerová, J. & Šilený, J. *Ann. Geophys.* **2**, 649–662 (1984).
4. Babuška, V., Plomerová, J. & Šilený, J. *Phys. Earth planet. Int.* (in the press).
5. Humphreys, E., Clayton, R. W. & Hager, B. H. *Geophys. Res. Lett.* **11**, 625–627 (1984).
6. Humphreys, E. D. & Clayton, R. W. *J. geophys. Res.* **95**, 19,725–19,746 (1990).
7. Babuška, V., Plomerová, J. & Granet, M. *Tectonophysics* **176**, 137–165 (1990).
8. Molnar, P. *Earth planet. Sci. Lett.* **101**, 68–77 (1990).
9. Woodward, R. L. & Masters, G. J. *geophys. Res.* **96**, 6351–6377 (1991).
10. Parmentier, E. M. & Phipps Morgan, J. *Nature* **348**, 325–328 (1990).

*Fifth International Conference on the Cell and Molecular Biology of *Chlamydomonas*, Pacific Grove, California, 26–31 May 1992.

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

FIG. 1 *Chlamydomonas asymmetrica*, seen in section and full false-colour glory. Chloroplasts are green; starch yellow; and the pyrenoid dark blue. Magnification is $\times 8,500$. (Science Photo Library.)

likely to be viewed as a form of recreation rather than an obstacle.

Successful cloning of the ASL gene by complementation of the *arg-7* auxotroph using a cosmid library was also welcome news, as it suggests that gene cloning by rather traditional approaches that take advantage of extant mutant collections is now possible (S. Purton, University College London). In addition, a number of new endogenous transposable elements, as well as the previously described *Gulliver*⁶, are being identified as common agents of spontaneous mutation, providing other tags for the identification and isolation of particular genes (J. Graham, Carnegie Mellon University; R. Schnell, University of Minnesota).

There was vicarious delight to be had in the report of heat-shock-induced transposon tagging of *Volvox* genes involved in nitrate metabolism, and the further characterization of *Jordan*, one of the mobile elements responsible (R. Schmitt, University of Regensburg). The successful cloning of the *Volvox* NR gene (by homology), and its use in obtaining a single transformation of the *nitA* host strain, is cause for hope that the surrender of this remarkable multicellular relative of *Chlamydomonas* may also be just around the corner. With transposon tagging of well-characterized developmental loci⁷ clearly now feasible, *Volvox* should become a powerful and popular model in molecular developmental biology.

Although nonhomologous integration of nuclear-transforming DNA remains problematic, the coexistence of engineered and resident wild-type alleles following nuclear transformation does not necessarily rule out analysis of the

effects of mutations on protein function. Cloned genes can be altered by site-directed mutation and then further tagged with unique epitopes. Expression of the transforming DNA then produces a mutant protein pool that can be tracked within cells by fluorescently labelled epitope-specific antibodies. This approach has been used to study the effect of tubulin acetylation on microtubule stability (K. Kozminski, Yale University). After first altering the normal acetylation site (the lysine-40 codon⁸), the cloned $\alpha 1$ -tubulin gene was tagged with an epitope from influenza virus haemagglutinin for which a monoclonal antibody was available. The mutant tubulin could then be

tracked in transformed cells by immunofluorescence and was found to assemble into both flagellar and cytoplasmic microtubules. Integration of the altered $\alpha 1$ -tubulin (presumably along with an equal number of wild-type $\alpha 1$ -tubulin monomers) had no apparent effect on microtubule function or microtubule stability, leaving the precise role of acetylation a mystery. Nevertheless, this approach should find wide application for studying the effect of specific mutations on the intracellular movement, compartmentalization and assembly of proteins.

Not surprisingly, work is underway to detect rare instances of homologous integration — and ultimately to increase their frequency — during nuclear gene transformation. A plasmid carrying only the 3' portion of the wild-type NR coding region has been used to transform a host deficient in nitrate reductase (O. Sodeinde, Cornell University). Correction of the host cell's defect could only occur through reversion of the resident mutation, or through homologous recombination between the intact resident copy and the introduced NR fragment (see Fig. 2). A unique insertion within an intron of the transforming NR fragment serves as a molecular footprint to distinguish between these events.

Linearization of the plasmid at a site within the NR fragment is essential for successful gene targeting, which occurs at frequencies two to three orders of magnitude lower than those of non-

homologous integration events in control transformations using the intact NR gene. But the relative frequency of targeted gene replacement is increased 40-fold when transformation is performed with the particle-bombardment system² rather than the glass-bead-vortexing method¹. The particle bombardment system (more commonly used for organelle transformations) delivers, on average, more DNA molecules per cell. Although this may in part account for the increased targeting of the transforming DNA, damage to the host chromosomal DNA by the DNA-coated tungsten particles may also act to stimulate the machinery for homologous recombination. The observation that different transformation protocols yield vastly different levels of homologous integration offers the prospect that conditions can be modified to further reduce the frequency of nonhomologous integrations when desired.

In a meeting of this sort, participants are united not by a common interest in a specific biological problem but by a common desire to know the organism better. The establishment of *Chlamydomonas* as an effective model for addressing a diversity of biological problems should become apparent over the months ahead as data from the more specialized ses-

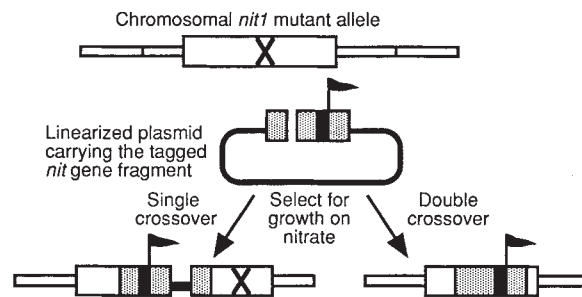


FIG. 2 Targeted gene replacement by homologous recombination in *Chlamydomonas*.

sions reach the journal pages. But it was messages issuing from the session on innovations in *Chlamydomonas* genetics and molecular biology, highlighted here, that will have the widest and deepest influence on the work of all *Chlamydomonas* biologists. □

Karen P. VanWinkle-Swift is in the Department of Biological Sciences, Northern Arizona University, PO Box 5640, Flagstaff, Arizona 86011-5640, USA.

- Kindle, K. L. *Proc. natn. Acad. Sci. U.S.A.* **87**, 1228–1232 (1990).
- Boynton, J. E. *Science* **240**, 1534–1538 (1988).
- Fernandez, E. et al. *Proc. natn. Acad. Sci. U.S.A.* **86**, 6449–6453 (1989).
- Debuchy, R. et al. *EMBO J.* **8**, 2803–2809 (1989).
- Choquet, Y. et al. *Cell* **52**, 903–913 (1988).
- Ferris, P. J. *Genetics* **122**, 363–378 (1989).
- Callahan, A. M. & Huskey, R. J. *Dev. Biol.* **80**, 419–435 (1980).
- LeDizet, M. & Piperno, G. *Proc. natn. Acad. Sci. U.S.A.* **84**, 5720–5724 (1987).

Lithium given and taken away

Gerry Gilmore

THE observation of substantial amounts of lithium in the vicinity of V404 Cygni, reported by Martín *et al.* on page 129 of this issue¹, will be of great interest to astrophysicists. There is roughly ten times as much lithium in the stars forming now as there is in stars that formed 15 billion years ago, when our Galaxy was relatively new. Where the additional lithium came from is one of the more enduring astrophysical puzzles. Various rare, energetic, short-lived sites are known to give lithium, but additional exotic sources have also been proposed. V404 Cygni, thought by many to contain a stellar-mass black hole, might be a distinctly unusual source.

The cosmic abundance of the element lithium is the result of competition between production and destruction. Destruction occurs through burning in stars, in the cores and in the high-density, million-degree region at the bottom of the deep convective zone which occurs naturally in the atmosphere of all stars similar to or cooler than the Sun. Considerately for astrophysicists, stars with surfaces somewhat warmer than the Sun's lack these deep convective zones, so their atmospheres retain the fossil record of the abundance of lithium in the gas from which the star formed. Thus, by analysis of suitable stars with a wide range of ages one can measure the evolution of the abundance of lithium.

More usefully, given the dubious nature of age calibrations in astrophysics, one can study the evolution of lithium in stars which show a wide range in their abundances of heavier elements, and thus truly represent the evolution of the chemical elements in the Galaxy. Such studies over the past few years have shown that the lithium abundance was more or less constant at a level near one part in 10^{10} of the hydrogen abundance for all stars with heavy-element abundances less than about one-tenth that of the Sun. The lithium abundance is then seen to rise roughly in proportion to the heavy-element abundance up to at least the solar abundance of heavy elements. Where does this extra lithium come from? And why didn't it become apparent at lower abundances?

Big Bang

Lithium was initially produced in the Big Bang, from which it emerged as the fifth most common element after ^1H , ^2H , ^3He and ^4He , with an abundance of about 10^{-10} that of ^1H according to the current standard Big Bang model. Or perhaps ten times this amount. Opinions differ. A respectable though not overwhelming

case can be made that the primordial abundance was close to the present high levels, and the tenfold decrease seen in metal-poor stars is a consequence of destruction even in those warm stars chosen for study. This hypothesis has received some impetus of late by precise analyses of metal-poor stars, which suggest a real diversity in the lithium abundances.

These analyses are nearing completion, and their publication is awaited with interest. Pending those studies, one may say that it is certainly not impossible that the lithium abundances in warm, metal-poor stars were initially higher than those seen today. Of course, it then remains a subject for further investigation to determine just how much higher, and where the extra lithium came from. It need not of course have been primordial, though even if it were the standard Big Bang model can cope. One simply requires a rather different amount of nuclear matter in the Universe.

Lithium is produced in various known sites in our Galaxy, and many other sites may be involved. One known source is cosmic-ray fusion, following the energetic collision of α -particles (helium nuclei) with one another. This process must have taken place to some degree when the Galaxy was still young, as helium nuclei were abundant from primordial production, and cosmic rays are accelerated in supernova remnants, which must have been plentiful then. The supernovae, and by implication the cosmic-ray flux, are known to have been plentiful, because we see and are made of their ashes — the heavy elements. Do we see the lithium created in the early Galaxy? The apparent near constancy of lithium in metal-poor stars may well be a false clue here, if some destruction has taken place in the way noted above. A simple self-consistent model of the early Galaxy, which also takes account of other recent work on the light element beryllium and its evolution in the early Galaxy, has the primordial lithium abundance perhaps 30–50 per cent below the present level in metal-poor stars, with the difference being cosmic-ray production. Again this lower primordial abundance is consistent with standard Big Bang models, and with the known production of lithium in other sites.

Stars an order of magnitude or so heavier than the Sun undergo a brief evolutionary state near the end of their lives known as hot-bottom convective-envelope burning. The physics is just that which allows the burning of lithium in lower-mass stars, though here with

precisely the opposite consequence. At the bottom of the convective envelope temperatures of 5×10^7 K or so are reached, allowing the ^7Be -producing reaction $^3\text{He}(\alpha, \gamma)^7\text{Be}$ to proceed. If rapid convective transport is available to remove the ^7Be from the hot environs before it is destroyed by proton capture, then its β -decay to give ^7Li can become a significant source of lithium. High lithium abundances have recently been observed in several stars in this evolutionary state, proving the significance of this reaction². Other known stellar sources include solar flares (this must be a minor contributor by Galactic standards), novae (possibly) and interacting close binary stars with extreme surface activity (RS CVn binaries)³.

Exotic sources

The more exotic possible sources of lithium include supernova shocks and photoerosion. This latter process, a succession of nucleon emissions from a heavy nucleus which can occur in a sufficiently intense field of high-energy radiation, is particularly intriguing. Recent calculations of the fate of moderately chemically evolved gas spending too long near to an active galactic nucleus⁴, and of primordial material near a massive black hole at redshifts of a few hundred⁵ (beyond the directly observable Universe), show that the light-element abundances can be very significantly modified.

The suggestion by Martín *et al.* in this issue¹, that lithium is seen to be produced in the environs of a stellar-mass black hole, is almost prosaic by comparison. There are several parts to their proposal, each worthy of consideration. Lithium certainly is seen in the spectrum of the stellar companion to the X-ray source V404 Cygni. It is demonstrated by Martín *et al.*; and in an independent and quite remarkable piece of astrophysics using just a plastic ruler, Wallerstein⁶ noticed the presence of the lithium absorption line in an earlier spectrum of V404 Cygni. The presence of lithium in V404 Cygni might not be too hard to understand, as spallation reactions of heavier elements in the hot, dense accretion disk around this putative black hole might well produce some lithium. How the newly produced lithium manages to get onto the black hole's companion star, which is supposedly losing mass to the accretion disk, remains something of a puzzle.

But is V404 Cygni really a black hole? In spite of extensive efforts, black hole searches have yet to provide a single really convincing candidate in the Galaxy. Perhaps this is the first. Whatever the outcome of further studies, and V404 Cygni is an interesting object worthy of much study, black holes are

not obviously common in the Galaxy. Their contribution to the lithium abundance is likely to remain small. It is more likely that the lithium abundance is a helpful clue towards an understanding of this specific object, rather than this object being a helpful clue towards an understanding of the evolution of lithium.

We are left with a picture of the formation and evolution of lithium in which the present abundance in warm stars with low heavy-element abundances may be a true record of primordial production, perhaps slightly modified by stellar depletion. Alternatively, much higher primordial production may have been compensated by greater depletion. Both these schemes may have to include significant production during the earliest stages of Galactic evolution. During the early stages of evolution of

the galactic disk, the lithium abundance rose by an order of magnitude. Or was perhaps depleted less. All this time, convection destroyed, hot-bottom convective burning created, cosmic rays created and modified, close binaries produced, and perhaps even black holes spalled — a remarkable cross-section of astrophysics, whose full understanding remains as an exercise for the reader. □

Gerry Gilmore is at the Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HE, UK.

1. Martin, E. L., Rebolo, R., Casares, J. & Charles, P. A. *Nature* **358**, 129–131 (1992).
2. Smith, V. V. & Lambert, D. L. *Astrophys. J.* **345**, L75–L78 (1989).
3. Pallavicini, R., Randich, S. & Giampapa, M. S. *Astr. Astrophys.* **253**, 185–198 (1992).
4. Boyd, R. N. & Fencel, H. S. *Astrophys. J.* **373**, 84–88 (1991).
5. Gnedin, N. Yu. & Ostriker, J. P. *Astrophys. J.* (submitted).
6. Wallerstein, G. *Nature* **356**, 569–570 (1992).

DOPAMINE RECEPTORS

Which D4 do you have?

Leslie Iversen

THE catecholamine dopamine continues to attract research interest because of its presumed pivotal role in brain pathways involved in mediating the actions of psychostimulant drugs and in maintaining 'drug seeking' behaviour for a variety of drugs of addiction. Dopaminergic mechanisms in the brain are also a key target for blockade by drugs used in the treatment of schizophrenic illness. On page 149 of this issue¹, Van Tol *et al.* show that one of the dopamine receptors in brain, the D4, exists in genetically determined polymorphic forms in the human population — this finding points to a possible basis for genetic differences in susceptibility to schizophrenia as well as for individual differences in response to drug treatment.

The history of biochemical studies of dopamine receptors extends over the past 20 years, following the discovery of Kebabian, Petzold and Greengard that some dopamine receptors in brain are positively coupled to adenylyl cyclase². Subsequent studies revealed that some of the dopamine antagonist neuroleptic drugs used to treat schizophrenic illness do indeed potently inhibit the dopamine-stimulated cyclase response, but that not all categories of clinically active neuroleptics are effective³. It became apparent that at least two subtypes of dopamine receptor existed: D1 sites, positively coupled to adenylyl cyclase, and the D2, which are not coupled or are negatively coupled⁴. Drug affinities to the D2 sites reflected most accurately their behavioural effects and their clinical potencies in treating psychosis. More

recently, both D1 and D2 receptors have been cloned^{5,6} and have been shown to belong to the G-protein-coupled family of receptor proteins. Within the past two years molecular genetic studies have revealed the existence of three further members of the dopamine receptor family — the D3 and D4 subtypes^{7,8}, which resemble D2, and the D5 subtype, which resembles more closely the D1 subtype⁹.

The D3 and D4 receptors resemble the D2 in their sensitivity to neuroleptic drugs, with some important differences. In particular, the D4 receptor is of considerable pharmacological interest as it binds the neuroleptic drug clozapine with an affinity some ten times higher than that of the D2 sites. This is intriguing because clozapine stands out clinically as one of the 'atypical' neuroleptics which retain antipsychotic efficacy but are much less likely to cause unwanted extrapyramidal motor side effects. This would probably have made clozapine the favoured drug for treating schizophrenic illness, were it not for the occurrence of rare haematological side effects. Could it be that the D4 receptor represents a site of especial importance in mediating the antipsychotic effects of such atypical neuroleptic drugs? Both D3 and D4 sites also have unusual anatomical distributions in brain; the messenger RNAs for these receptors are expressed in highest densities in limbic forebrain areas, thought to be key targets for antipsychotic drug actions, rather than in the basal ganglia which mediate extrapyramidal side effects.

Van Tol *et al.*¹ now add a new dimension to the complexity of dopamine pharmacology with their discovery that there are different polymorphic forms of the D4 receptor in the population. Screening human genomic libraries revealed three different forms, containing, respectively, two-, four- or sevenfold repeats of a 48-base-pair sequence in the putative third cytoplasmic loop region. A further two forms may exist containing one or five repeats of this sequence. These differences persist in the expressed receptors and, furthermore, the different forms of the D4 receptor display pharmacological differences when expressed in COS cells; the seven-repeat form, for example, is less sensitive to changes in agonist or antagonist (clozapine) affinity induced by addition of sodium chloride.

Whether the pharmacological differences observed *in vitro* translate into meaningful differences in drug sensitivity in the intact brain remains to be seen. The identification of these unexpected human variants of the D4 receptor, however, opens up many new and attractive vistas for molecular genetic studies of psychiatric illnesses¹⁰. Is it possible that the genetic disposition to schizophrenic illness could be linked to the presence of particular subtypes of dopamine receptor in susceptible individuals? Might the same be the case in those more likely to become addicted to drugs, which act in part through central dopaminergic systems?

In this fast-moving field there are inevitably more questions than answers. What, for instance, are we to make of the recent report by O'Malley *et al.*¹¹, claiming that the rat equivalent of the dopamine D4 gene is preferentially expressed in cardiovascular tissues? Are there yet more dopamine receptor subtypes to be discovered? Why have so many evolved? And, perhaps of most general interest, will this new knowledge help in the design of improved drug treatments for psychotic illness or drug addiction? □

Leslie Iversen is at the Merck Sharp & Dohme Research Laboratories, Neuroscience Research Centre, PO Box 1675, Harlow, Essex CM20 2PT, UK.

1. Van Tol, H. H. M. *et al.* *Nature* **358**, 149–152 (1992).
2. Kebabian, J. W., Petzold, G. L. & Greengard, P. *Proc. natn. Acad. Sci. U.S.A.* **79**, 2145–2149 (1972).
3. Iversen, L. L. *Science* **188**, 1084–1089 (1975).
4. Kebabian, J. W. & Calne, D. B. *Nature* **277**, 93–96 (1979).
5. Deary, A. *et al.* *Nature* **347**, 72–76 (1990).
6. Bunzow, J. R. *et al.* *Nature* **336**, 783–787 (1988).
7. Sokoloff, P., Giros, B., Martres, M. P., Bouthenet, M. L. & Schwartz, J.-C. *Nature* **347**, 146–151 (1990).
8. Van Tol, H. H. M. *Nature* **350**, 610–614 (1991).
9. Sunhara, R. K. *et al.* *Nature* **350**, 614–619 (1991).
10. Mendlewicz, J. & Hippus, H. (eds) *Genetic Research in Psychiatry* (Springer, Berlin, 1992).
11. O'Malley, K. L., Hermon, S., Tang, L. & Todd, R. D. *New Biol.* **4**, 137–146 (1992).

RÉSUMÉ

Visible benefit

WHEN trichromatic colour vision has such obvious evolutionary advantages, why should some 2 per cent of males in a human population be dichromats (that is, unable to distinguish between red and green)? The question is posed by M. J. Morgan *et al.* (*Proc. R. Soc. B* **248**, 291–295; 1992), who return to the idea mooted during the Second World War that dichromats can 'see through' camouflage that foxes the colour-normal observer. In simple but ingenious tests, colour was used to mask targets demarcated from a background by the orientation or size of their texture elements. Dichromats were indeed significantly better at spotting the camouflaged region. But although this ability in disability is undoubted, the authors allow that the polymorphism of colour vision may be maintained purely by the genetic mechanism of unequal crossing-over.

In season

SATURN's moon Titan, the second largest satellite in the Solar System and the only one with an appreciable atmosphere, has seasons, according to new observations from the Hubble Space Telescope (J. Caldwell *et al.* *Icarus* **96**, 1–9; 1992). In 1990, when the observations were made, Titan's Northern Hemisphere was brighter than its Southern Hemisphere, the reverse of what Voyager 2 saw in 1981, a third of a Titan year earlier. The role of 'seasons' in modulating Titan's atmosphere and causing slow variations in brightness and the North–South asymmetry was suspected in 1981, but a competing role for the 11-year solar cycle was also implicated. Related observations show that what changes you see depend on the wavelength you look at, as different parts of the spectrum reveal different parts of Titan's atmosphere.

Current communication

Gymnotus carapo is territorial, and can distinguish between its neighbours' voices and associate them with the appropriate territories. *G. carapo* is not a songbird — it is a fish and uses electrical pulses for its 'song' (P. K. McGregor and G. W. M. Westby, *Animal Behaviour* **43**, 977–986; 1992). The pulsed signals are separated by timed silences or 'interpulses', whose lengths are inversely proportional to dominance status. Yet even with interpulses removed from playbacks, fish respond in characteristic ways, because the waveforms are as characteristic of individual fish as songs are of birds. The brevity of each pulse (just two μ s) is close to the speed of nervous transmission, however, so it is hard to understand how the fish can tell between any two — it could be that they respond to the 'beat' frequency differences against their own signal.

INFLUENZA VIRUS

Amantadine blocks the channel

J. J. Skehel

THE membrane components of viruses such as influenza encounter acidic cellular compartments at two stages of virus replication. At the beginning of infection viruses are taken into endosomes; there their fusion glycoproteins are activated at low pH to fuse virus and endosomal membranes, allowing the genome–transcriptase complex to enter the cell. Later, newly synthesized fusion glycoproteins and other virus membrane proteins are transported to the cell surface through the acidic *trans*-Golgi. It is thought that in both of these environments one of the influenza virus membrane proteins (M_2 , which consists of 97 amino acids) functions as an ion channel, and this property of the protein is addressed by Pinto *et al.* in *Cell*¹.

Using voltage-clamp procedures to analyse the total membrane currents of oocytes injected with M_2 messenger RNA, the authors show that expression of M_2 at the cell surface correlates with the activity of a novel ion channel, and that the current carried by the channel is regulated at low pH. They also report that channel activity is blocked by the anti-influenza drug amantadine, and that the channel properties of mutant M_2 proteins (which have been reported to confer resistance to amantadine during virus infections) are unchanged in amantadine-treated oocytes. Studies of such mutants and of the mechanisms of action of amantadine have formed the basis for proposals that, during virus replication, M_2 functions as a channel in two stages — initially as a virus membrane component, which allows acidification of the virus core; and subsequently as a *trans*-Golgi membrane component, which relieves the pH gradient in this compartment^{2,3}.

The anti-influenza activity of amantadine (1-amino adamantane hydrochloride) was first described in 1964 (ref. 4). Since then, it has been shown to be effective in clinical trials⁵ but has had limited general use, for two main reasons. One is that it can cause neurological side effects⁶. The other is that its high specificity for influenza A viruses and lack of effect on influenza B (ref. 4) requires identification of the virus concerned before correct prescription. From the time of the first reports of its activity, it has been known that addition of amantadine to cells before infection blocks an early event in influenza replication. The possibility was therefore considered that, as a weak base, it increases endosomal pH (ref. 7) and prevents the low pH-activation of the membrane fusion potential of the haemagglu-

tinin fusion glycoprotein. The isolation of amantadine-resistant mutants containing haemagglutinins which fused membranes at higher pH than wild-type virus was consistent with this possibility⁸.

However, several studies^{9,10} have shown that at much lower concentrations, which do not influence endosomal pH, amantadine prevents the disassembly of the virus core. As a consequence it blocks virion transcriptase activity and inhibits its transfer to the nucleus where virus mRNA synthesis normally occurs. Virus mutants resistant to this blockade all contain mutant M_2 proteins with amino-acid substitutions in their transmembrane regions¹¹. Amantadine can also prevent virus replication at a later stage, and it is primarily from studies of this event that the proposed ion-channel function of M_2 is derived¹².

In this case, treatment of infected cells with amantadine leads to the detection of haemagglutinin molecules in the *trans*-Golgi and at the surface of infected cells in a conformation characteristic of the conformation they are induced to form at the low pH of virus–membrane fusion. Viruses containing a mutant haemagglutinin which is more stable than wild type, and which is insensitive to premature activation at *trans*-Golgi pH, are resistant to this effect of the drug¹³. But more importantly, the same mutations in M_2 which override the early block in replication also lead to drug resistance¹².

Three conclusions have been drawn from these results. First, that by blocking M_2 function amantadine maintains the *trans*-Golgi at its normally low pH; second, that the initial inhibition of virion core disassembly results from blockage of the same function; and third, that M_2 functions as an ion channel in both cases, mediating the transmembrane export of H^+ from the *trans*-Golgi and the import of H^+ to the interior of infecting virus particles in endosomes.

The observations of Pinto *et al.*¹ — that amantadine blocks the ion-channel activity of wild-type M_2 in oocytes and is without effect on the channel properties of the M_2 mutants — provide strong support for these conclusions. In addition particular attention is given to the question of whether M_2 functions independently as a channel, as might be expected if the proposal that it functions as a virion channel component is correct, or whether it in some way modifies the properties of an endogenous channel. Using two mutant M_2 proteins with either an additional amino acid or a deletion of four amino acids in the transmembrane region, which they pre-

sume are without effect on the overall structure of the channel, the authors observe that, unlike wild-type M_2 , mutant channel activation is voltage dependent and this property can be used to show that the mutants differ in selectivity for both Na^+ and Cl^- . The different permeabilities suggest that the mutations directly affect the pore-forming region of the channels, which clearly implies that the transmembrane region of M_2 forms the pore and that M_2 itself functions as a channel.

Similar experiments and rationales were recently used¹⁴ in the characterization of the single-transmembrane subunit, MinK, K^+ -selective channels found, for example, in mammalian kidney¹⁵. Perhaps influenza M_2 , because of its similar size and subunit structure, may offer similar advantages for investigations of the basic mechanism of channel function. And perhaps direct investigation of the mechanism of amantadine action, which the studies of Pinto *et al.* will allow particularly if extended to single-channel recording of purified M_2 reconstituted into lipid bilayers, may direct the search for other antiviral chan-

nel blockers. They will certainly stimulate the examination of other virus systems for proteins with similar functions, possibly including influenza B, and may even lead to a generally applicable anti-influenza prophylactic. □

J. J. Skehel is at the MRC National Institute for Medical Research, Mill Hill, London NW7 1AA, UK.

1. Pinto, L. H., Holsinger, L. J. & Lamb, R. A. *Cell* **69**, 517–528 (1992).
2. Hay, A. J. in *Concepts in Viral Pathogenesis III* (eds Notkins, A. L. & Oldstone, M. B. A.) 561–567 (Springer, New York, 1989).
3. Hay, A. J. *Sem. Virol.* (in the press).
4. Davies, W. L. *et al. Science* **144**, 862–863 (1964).
5. Dolin, R. *et al. New Engl. J. Med.* **307**, 580–584 (1982).
6. Bryson, Y. J., Monahan, C., Pollack, M. & Shields, W. D. *J. Infect. Dis.* **141**, 543–547 (1980).
7. Marsh, M. & Helenius, A. *Adv. Virus Res.* **36**, 107–151 (1989).
8. Daniels, R. S. *et al. Cell* **40**, 431–439 (1985).
9. Bukrinskaya, A. G., Vorkunova, N. K. & Pusharskaya, N. L. *J. gen. Virol.* **60**, 61–66 (1982).
10. Martin, K. & Helenius, A. *Cell* **67**, 117–130 (1991).
11. Hay, A. J. *et al. EMBO J.* **4**, 3021–3024 (1985).
12. Sugrue, D. A. *et al. EMBO J.* **9**, 3469–3476 (1990).
13. Steinhauer, D. A. *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 11525–11529 (1991).
14. Goldstein, S. A. N. & Miller, C. *Neuron* **7**, 403–408 (1991).
15. Hausdorff, S. F. *et al. Biochemistry* **30**, 3341–3346 (1991).

VERTEBRATE DEVELOPMENT

Induction and all that

Jonathan Cooke

DEVELOPMENTAL biology has seemed always to alternate between periods of apparent clarification and obfuscation, and the undoubted breakthrough into its truly molecular era has not yet carried the field beyond this pattern. Indeed to those who relish conundrums, therein lies much of its charm. So much was clear at a conference* last month, the vast remit of which included the latest findings concerning the 'inductive' intercellular signalling system that sets up the axial plan of the frog's body. With their large, rapidly developing eggs, accessible to micromanipulation as well as molecular biology, frogs should have a key role in elucidation of this (conserved, one hopes) system in vertebrates.

Just two years ago all had seemed rather clear in outline (if only to *naïfs* such as myself), and that outline is shown in the figure, overleaf. An identical baseline or 'default' state that would lead only to epidermal differentiation seems to be inherited by all cells of the animal or ectodermal hemisphere of the cleaving blastula (frog embryos do not truly grow during the stages of pattern formation). Two sharply distinct new states are initially induced, respectively in minor and major sectors of the

equatorial region, by signals emanating from cells of the yolk vegetal hemisphere that itself has no long-term future in body structure. The body plan then emerges by subsequent interactions between this pair of territories that are induced first — the dorsal-axial and ventrolateral marginal zones (DMZ and VMZ) — and between these zones and the remaining animal cap ectoderm.

Three contributions at the meeting revealed strikingly that in the presumptive nervous system, at the poleward edge of the DMZ (see figure), distinctive spatially patterned cell-mechanical activity (R. Keller *et al.*, University of California, Berkeley) and gene activity (T. Doniach, University of California, Berkeley; L. Essex, NIMR, London) are initiated as part of the one episode of inductive 'planar' signalling, within the blastula and early gastrula wall, that also founds the meso-endodermal pattern. Thus in the few hours available, and within the plane of a truly cellular tissue, spatial complexity must become comparable with that achieved by the famed *Drosophila* blastoderm. But in the vertebrate this process must involve extensive deployment of truly (if only locally) intercellular signals, probably of classes yet uncharacterized.

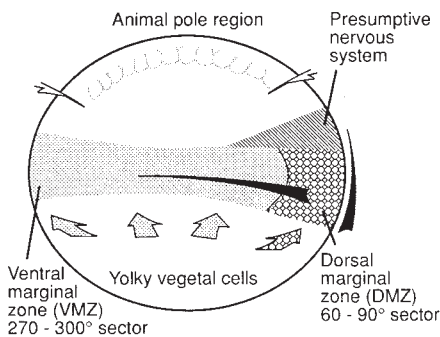
The signals initiating DMZ and VMZ

specifications had seemed securely characterized as members of the activin and the fibroblast growth factor (FGF) peptide families, respectively, short of precise identification of the molecules involved *in vivo*^{1–4}. But while the primacy of roles for members of these families becomes more probable with each appropriate experiment, the nature of those roles becomes less clear. Thus the long-awaited functional knockout of the activin cell-membrane receptor, by loading the egg with RNA for a dominant negative mutant subunit (A. Hemmati-Brivalou, communicated by M. Ku, Harvard University), can give a phenotype that is mesodermless (lacking apparent induction) on all meridians — not just without a DMZ to control axial pattern, as the earlier scheme would predict. Reciprocally, though more equivocally, the corresponding knockout phenotype specific to the FGF receptor class suggests progressive disruption, rather than obliteration, of induction on embryo meridians fated for more posterior axial regions (E. Amaya, University of California, San Francisco), instead of the massive deficit in all but the restricted DMZ sector that 'the model' would predict. This implies that activin is needed for initial induction over most of the fate map of the normal embryo, whereas FGF has more of a modulatory effect on the positional (head-tail or dorsal-ventral) character of what is induced, being progressively more important in more posterior body regions.

One interpretation of such observations is that, at least in this sort of vertebrate development, it is a spatially graded conversion of activin to its active form rather than its synthesis in one place which is the pattern-organizing principle, against a relatively ubiquitous background presence of an FGF-like inducing component. This would render more comprehensible the observations that an activin-like component is supplied maternally as protein, apparently around much of the growing *Xenopus* oocyte (G. Guex, NIMR, London), while at least one FGF component is widespread at blastula stages inside marginal zone cells, over much of what would be thought of as the induced rather than the signal-synthesizing or signal-storage region (S. Godsave, Furosawa Morphogene Project).

When activin is caused to be secreted locally within a blastula by loading one of its large, early cleavage cells with the injected RNA, a properly organized but anteriorly incomplete body plan is induced in relation to that site. The incompleteness of that induced pattern is unhesitatingly ascribed by molecular biologists to use of 'not quite the right' or 'not the only necessary' *in vivo* localized molecule. But might not the im-

* Fourth *Xenopus* International Conference, Pacific Grove, California, 4–7 June 1992.



Schematic side view of the *Xenopus* embryo at late blastular stage, with the future dorsal midline to the right. Shaded areas across the equatorial region represent the first-induced sectors referred to in the text, while the matching arrows beneath represent the belief that two specific arrays of vegetal-inducing signals are responsible for induction. The long black arrowheads represent the idea that signalling systems for further organization could relate to the progressive distance from the first signal sources within the marginal zone, and the function of DMZ as a new source for signalling into VMZ. Both of these dimensions tend to map onto the head → tail organization of the future body. Also indicated is how the cell sheet forming the animal pole region can be excised for culture in a simple medium with purified proteins (or without them, if whole eggs have been pre-loaded with an ectopically transcribable RNA) to test candidate developmental signalling molecules. It can be seen that the extent of animal pole taken, and the age of the donor blastula, could affect the interpretation of such tests.

perfection stem from the spatial crudity of this experiment, in relation to some pattern of source localization or signal timing that attends normal development? That speculation remains worthwhile in my view, but the reductionist approach may nevertheless have been vindicated by the identification of two powerful new players whose injection in this way does induce head along with the remaining dorsal axial pattern.

The Wnt proteins (the founder member of which was the oncogene *int-1*, in turn cognate with the *Drosophila* segment polarity gene *wingless*) might seem good candidates for pattern organization in embryos. They are matrix-associated on secretion so that their effects might spread only slowly and across some six cell diameters⁵. Most though not all members thus far cloned from *Xenopus* or mammals — as well as the *Drosophila* homologue — are effective in achieving head pattern in blastomere injection or comparable experiments (L. Christian, Washington University, Seattle; L. Dale, Birmingham University). One such effective RNA is present at the right vegetal location and times to be a candidate in normal development (M. Ku). Noggin (R. Harland, University of California, Berkeley), a more conventionally secreted protein but one with no known

relatives to date, cloned directly from *Xenopus*, has very similar properties to the Wnt proteins in RNA-injection experiments.

On balance, Noggin and Wnt proteins are deemed themselves not to be inducers (that is, able to effect a change in specified state in naive animal cap cells *in vitro*, like activins and FGFs). Rather they may modulate the effects of inducers locally, leading to enhanced DMZ states that allow development of the anteriormost body parts. This raises the possibility that inducers as such play less critical spatial roles in the organization, while other adjunct signals, in modulating the local character of induced states, adopt those roles — for example Noggin and/or a Wnt protein as 'dorsalizers' (R. Harland). Thus the animal hemisphere from early stages, before overt induction, may become cryptically patterned or graded in responsiveness by the regional influence of such molecules from more vegetal cells (or, conceivably, by direct inheritance from egg structure).

Evidence of such cryptic pre-organization of the 'uninduced' animal hemisphere has been controversial since before the molecular breakthrough. Modulators might, after all, be synthesized as immediate downstream consequences of induction rather than being received as signals ahead of time. At the meeting, this controversy took the form of contention over whether animal hemispheres, of particular ages, are or are not subregionalized in the local responsiveness of their cells to activin *in vitro*. Laboratories separated merely by the Atlantic Ocean differ as to whether groups of animal hemisphere cells from ventral sides of blastulae, cultured in a simple salt solution with particular concentrations of the protein, can give rise to the unambiguous tissue notochord as dorsal animal cells are agreed to do.

While such episodes recur, the field remains a rich store of human as well as intellectual interest. But, quite properly and indignantly, grass roots postdocs and students ask why lab heads cannot spare the jettare that would enable pairs of young, clever and open-minded people to sit down side by side with their tungsten needles and activin samples, and find out what is going on. □

Jonathan Cooke is at the MRC National Institute for Medical Research, Mill Hill, London NW7 1AA, UK.

1. Smith, J. C., Price, B. M. J., Van Nimmen, K. & Huylebroek, D. *Nature* **345**, 729–731 (1990).
2. Smith, J. C., Cooke, J., Green, J. B. A., Howes, G. & Symes, K. *Development* **107**, Suppl. 149–159 (1989).
3. Slack, J. M. W., Darlington, B. G., Heath, J. K. & Godsave, S. F. *Nature* **326**, 197–200 (1987).
4. Kimelman, D., Abraham, J. A., Haaparanta, T., Palisi, T. M. & Kirschner, M. W. *Science* **242**, 1053–1056 (1988).
5. Jue, S. F., Rogers, S. B., Rudnick, J. A., Varmus, H. E. & Brown, A. M. C. *Molec. cell. Biol.* **12**, 321–328 (1991).

A light grip

MODERN high-performance cars are limited by the fundamental laws of friction. All the forces acting on the car come from the reaction of the tyres on the road; even with unity coefficient of friction these forces cannot exceed the weight of the car. Four-wheel drive, antilock braking systems, power steering and so on all try to get as close to this limit as possible. Racing cars try to exceed it: they use 'negative lift' surfaces which push the car down onto the road and increase its effective weight. But what we really need, says Daedalus, is a sticky tyre.

At first sight, this seems obviously unworkable. Conventional repeated-stick surfaces — such as office labels and adhesive putties — absorb a great deal of energy during each stick–unstick cycle. A tyre so coated would have enormous drag. But Daedalus recalls the chemistry of photodecomposition. Light excites the electrons of a chemical bond into some antibonding orbital. The bound atoms are not merely uncoupled; they are actively pushed apart. In the dark, they can recombine again. It all happens in picoseconds.

So DREADCO chemists are devising a photoactive tyre surface. Its marginally stable molecules are easily disrupted by daylight; its illuminated surface is populated largely by antibonding orbitals. But as the tyre revolves, each section in turn is pushed against the road. The daylight is cut off, and the photoactivated free radical surface reverts to a bonding state. It reacts with the road surface, gluing itself firmly in place. Meanwhile, at the trailing edge of the tyre, light leaks in from behind. It continuously excites the emerging surface back to the antibonding state; it unsticks instantly, and is positively repelled from the road.

Thus four-wheel drive, antilock braking and so on will be rendered obsolete. Fitted with DREADCO's phototires, any car will enjoy utterly tenacious roadholding — at least in daylight. At nightfall, of course, it will grind to a halt, glued irretrievably to the road until the next day. Some car owners would welcome this defence against nocturnal theft; but for those who wish to drive at night, lamps could be mounted under the wheel-arches to shine on the tyres.

DREADCO's photoadhesive technology should find other markets too. Photoboots for runners and mountaineers would transform these sports, while photogloves would give a grip of iron to pole vaulters, steeplejacks and hammer welders. Both would be welcomed by mariners on storm-tossed barques and bus conductors on hilly or ill-maintained routes.

David Jones

Proto-vav and gene expression

SIR — Two recent papers^{1,2} describe the *vav* proto-oncogene product as containing 'leucine-zipper' (LZ) and 'zinc-finger' (ZF) motifs, two nuclear localization signals and a 'helix-loop-helix' (HLH) domain. We were surprised to find, however, that none of these transcription-factor and DNA-binding motifs attributed to proto-*vav* is detectable when objective or quantitative criteria are applied.

We examined the proto-*vav* amino-acid sequence as follows. Initial database searches (see figure legend) confirmed the presence of two SH3 and one SH2 domains but failed to provide evidence of sequence similarity to any ZF, HLH or LZ proteins. In contrast, the relationships between *max*, *myc* and other HLH/LZ proteins were easily detectable and highly significant ($P < 10^{-7}$) under the same search conditions. Considering perhaps that similarities between proto-*vav* and other HLH/LZ and ZF sequences might be obscured by the more significant search results, we performed subsequent analyses with an edited query sequence encompassing residues 1–613 of proto-*vav* but excluding the C-terminal SH3–SH2–SH3 domains. A search of the general-purpose sequence databases revealed significant local similarities between proto-*vav* (residues 513–561) and the diacylglycerol/phorbol ester (DAG/PE) binding domains of protein

kinase C — a fact recognized by others (for example, ref. 3) and documented in the PROSITE database (see below). Proto-*vav* also shows some borderline similarities to the *dbl* oncogene product⁴ and some insignificant local similarities to intermediate-filament proteins (see below).

Directed searches of the Zinc Finger⁵ and Transcription Factor⁶ databases provided no evidence for even marginal similarity between *vav* and any of the 1,313 ZF domains and 811 transcription-factor domains present in these collections. Finally, the complete proto-*vav* sequence was scanned for the presence of any of the 605 motifs in the PROSITE database⁷. Proto-*vav* does not meet the criteria for nuclear localization signals and the 'zinc-finger' region corresponds to the DAG/PE-binding motif mentioned above (PROSITE entry PS00479). Statistical methods confirmed the presence of an acidic domain (residues 130–174) in proto-*vav*, but charge clusters associated with DNA-binding proteins and transcription factors are usually basic in nature⁸.

Adams *et al.*⁹ have just published a revision of the mouse proto-*vav* sequence affecting 32 codons specifying residues 325–355 of the protein. When this revised sequence is used in database searching as described above, the results are essentially the same as before with

the profound exception that the sequence now shows extremely significant ($P \sim 2 \times 10^{-10}$) similarity to a family of proteins (including human Dbl, Bcr and yeast Cdc24) that act as guanine nucleotide dissociation stimulators for the *rho/rac* family of *ras*-like small GTPases (see ref. 10). This strongly confirms the results of Adams *et al.*⁹ and supports their conclusion that *vav* functions in a signal transduction pathway involved in cytoskeletal organization.

Because of the local similarities between proto-*vav* and some intermediate-filament proteins, we determined the propensity of the proto-*vav* sequence to adopt a coiled-coil conformation. We found three regions (residues 148–178, 281–311, 351–379) with coiled-coil probabilities of about 50% (*b* in the figure). Note that these regions do not include the putative leucine-zipper motif (residues 70–91). A summary of some objective and quantitative sequence attributes of proto-*vav* is shown in *a* in the figure. None of our findings supports the hypothesis^{1,2} that proto-*vav* might have DNA-binding or transcription-factor activities.

Mark S. Boguski

National Center for Biotechnology Information,

National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894, USA

Amos Bairoch

Department of Medical Biochemistry, University of Geneva, 1211 Geneva 4, Switzerland

T. K. Attwood

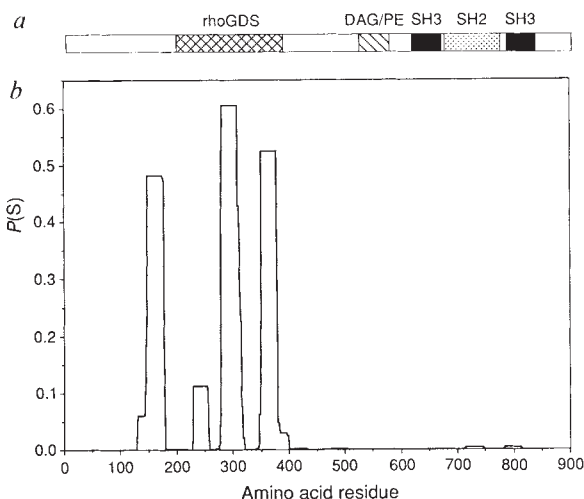
Department of Biochemistry and Molecular Biology, University of Leeds, Leeds LS2 9JT, UK

George S. Michaels

Division of Computer Research and Technology, National Institutes of Health, Bethesda, Maryland 20892, USA

Computable sequence attributes of the *vav* proto-oncogene product. The complete human proto-*vav* sequence was constructed by merging 5' data¹¹ with the SWISS-PROT entry for *vav* (accession no. p15498), excluding the Tn5-derived residues at the amino terminus of the latter sequence. Proto-*vav* was then searched against a merged, non-redundant collection of sequences derived from PIR 31.0, SWISS-PROT 21.0 and translated GenBank 71.0 using the BLASTP program¹² with default parameters. Searches against the Zinc-Finger and Transcription-Factor databases were conducted with FASTA¹³ and BLASTP¹², respectively.

This analysis was repeated with the revised mouse proto-*vav* sequence^{3,9}. *a*, Schematic summary of significant search results as described in the text. The sizes and locations of the rhoGDS, DAG/PE and SH3, SH2 and SH3 domains are drawn to scale and correspond to residues 198–385, 516–564, 617–665, 671–665 and 789–837, respectively, in the mouse proto-*vav* sequence. *b*, The proto-*vav* sequence was also analysed for the presence of potential coiled coil regions using the method of Lupas *et al.*¹⁴ with a window of 28 residues. The probabilities of segments forming coiled coils are plotted as a function of their location in the linear sequence. The regions showing some coiled-coil potential correspond to the rhoGDS domain (previously described as the leucine-rich domain³) but do not include the putative leucine-zipper region. *P*(S) is the probability of forming a coiled coil of score *S* defined in ref. 14.



1. Bustelo, X. R., Ledbetter, J. A. & Barbacid, M. *Nature* **356**, 69–71 (1992).
2. Margolis, B. *et al.* *Nature* **356**, 71–74 (1992).
3. Coppola, J., Bryant, S., Koda, T., Conway, D. & Barbacid, M. *Cell Growth Diff.* **2**, 95–105 (1991).
4. Galland, F., Katav, S. & Birnbaum, D. *Oncogene* **7**, 585–587 (1992).
5. Jacobs, G. & Michaels, G. *New Biol.* **2**, 583–584 (1990).
6. Ghosh, D. *Trends biochem. Sci.* **16**, 445–447 (1991).
7. Bairoch, A. *Nucleic Acids Res.* **19**, 2241–2245 (1991).
8. Brendel, V. & Karlin, S. *Proc. natn. Acad. Sci. U.S.A.* **86**, 5698–5702 (1989).
9. Adams, J. M., Houston, H., Allen, J., Lints, T. & Harvey, R. *Oncogene* **7**, 611–618 (1992).
10. Boguski, M. S. & McCormick, F. *Neuron* (in the press).
11. Katav, S., Martin-Zanca, D. & Barbacid, M. *EMBO J.* **8**, 2283–2290 (1989).
12. Atschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. *J. molec. Biol.* **215**, 403–410 (1990).
13. Pearson, W. R. & Lipman, D. J. *Proc. natn. Acad. Sci. U.S.A.* **85**, 2444–2448 (1988).
14. Lupas, A., Van Dyke, M. & Stock, J. *Science* **252**, 1162–1164 (1991).

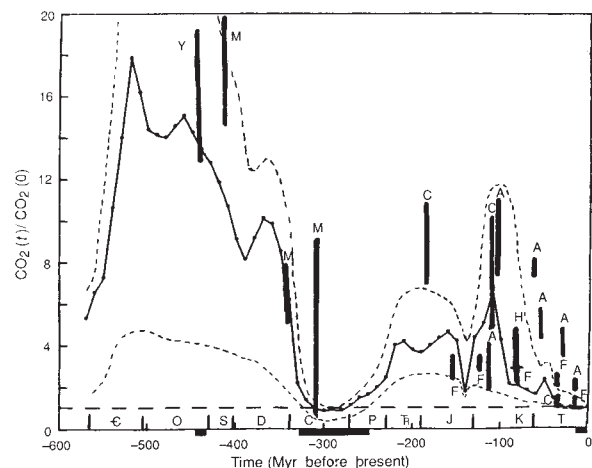
Palaeo-CO₂ and climate

SIR — As recognized by Kasting in his News and Views article¹, there is a problem in reconciling high estimated atmospheric CO₂ levels^{2,3} for the late Ordovician period (about 440 million years (Myr) ago) with the occurrence of continental glaciation at that time. Nevertheless, other estimates of palaeolevels of CO₂ are in general agreement with independent estimates of palaeoclimate for the remaining portion of Phanerozoic time (the past 570 Myr). The two most thoroughly studied Phanerozoic glaciations are those of the Permo-Carboniferous and late Cenozoic which occur at times of minimum CO₂ (see figure). Also, well-documented warm periods, such as the Cretaceous (symbol K, 135–65 Myr ago), correspond to high estimated levels of atmospheric CO₂. Taken together, these results all point to the importance of the atmospheric greenhouse effect as a major control on very long-term climate.

The methods used in constructing the figure are quite different. The curve is the best estimate obtained from a computer model of the long-term carbon cycle^{3,4}. This model considers the effect on atmospheric CO₂ of the large and dominant carbon fluxes over millions of years between rocks and the atmosphere/hydrosphere/biosphere system. This includes: (1) weathering of Ca and Mg silicates and carbonates and sedimentary organic matter on the continents; (2) the burial of carbonates and organic matter in sediments; and (3) the thermal breakdown of carbonates and organic matter at depth with resultant CO₂ degassing to the atmosphere. The area, mean elevation and position of the continents, the rise and evolution of vascular land plants, and the feedback effect of CO₂ on both greenhouse global warming and plant growth are considered as important factors affecting weathering rates. The carbon isotope composition and abundance of carbonates and organic matter in sedimentary rocks are used to calculate burial rates (see also ref. 5). Rates of seafloor creation and subduction, as revealed by plate-tectonic theory and changes in sea level, are assumed to be a direct measure of degassing rate. Many of these parameters are not well known and variation of their values between reasonable limits gives rise to the range of CO₂ values enclosed by the dashed lines in the figure. Also, this type of modelling is subject to constant revision as better methods of estimating the various parameters arise.

Another method for calculating palaeo-CO₂ is the measurement of the

Carbon dioxide against time based on a theoretical model of the long-term carbon cycle^{3,4}. The parameter CO₂(t)/CO₂(0) is the ratio of mass of CO₂ in the atmosphere at some time in the past to that at present. Dashed lines enclose the approximate range of error of the model based on sensitivity analysis⁴. Horizontal bars along the abscissa, major glaciations; vertical bars, independent palaeo-CO₂ estimates based on: (1) carbon isotope analyses of CaCO₃ in palaeosols (C, ref. 6; M, ref. 7); (2) carbon content and isotope composition of the FeCO₃ component of goethite in palaeosols (Y, ref. 2); (3) the difference in carbon isotope composition between organic matter and CaCO₃ in marine sedimentary rocks (A, ref. 8; H, ref. 10); and (4) the difference in carbon isotope composition between specific biomarker compounds and CaCO₃ in marine sedimentary rocks (F, ref. 9). Lengths of vertical bars represent each published estimate of range in error. (The 320 Myr value is corrected for change in the carbon isotope composition of the ocean and atmosphere and the probably arid character of the sample⁷.)



¹³C content of CaCO₃ (refs 6,7) and the FeCO₃ component of goethite² in ancient palaeosols. Both methods assume that the ¹³C content of pedogenic carbonate represents a mixture of ¹³C-enriched atmospheric CO₂ and ¹³C-depleted CO₂ generated from the decomposition of soil organic debris. The palaeolevel of CO₂ at depth in the soil is assumed constant in the CaCO₃ method, whereas in the other method it is calculated from the FeCO₃ content of the goethite. There are many assumptions in these methods and the range of error stated in each study referenced is shown in the figure.

A third method of palaeo-CO₂ estimation is based on the fractionation of ¹³C between carbonates and organic matter in marine sedimentary rocks^{8–10}. The isotope composition of carbonate carbon reflects ancient seawater composition and the difference between this and the isotope composition of organic carbon reflects the degree of carbon isotope fractionation during photosynthesis. This fractionation has been shown to be a function of the CO₂ content of sea water^{10–12}, so that assuming equilibrium with the atmosphere, one can calculate levels of palaeo-CO₂. Lack of ocean-atmosphere equilibration, vital effects on carbon-isotope fractionation, differences in temperature, and diagenetic processing of organic matter can all give rise to errors in this method. (To avoid the last problem the ¹³C content of unaltered biomarker compounds only are analysed in ref. 9.) This helps explain the length of each vertical bar and the disagreement between studies shown in the figure.

Regardless of the numerous errors mentioned above, there is an amazing semi-quantitative agreement between the results of the various methods for

estimating palaeo-CO₂. In other words, the times of high versus low CO₂ are similar and actual calculated values agree within a factor of about three. Kasting's hope¹ for methods of deducing ancient CO₂ levels as accurate as those based on the study of trapped air bubbles in glacial ice is probably overly optimistic because of various inherent errors in the carbon isotope methods, as well as large ones in the model calculations. Nevertheless, the isotope methods do hold promise for making semi-quantitative estimates of ancient CO₂ levels, and modelling calculations at least point to important processes affecting CO₂ and the need for a better understanding of them.

Robert A. Berner

Department of Geology and Geophysics,
Yale University,
New Haven,
Connecticut 06511, USA

1. Kasting, J. F. *Nature* **355**, 676–677 (1992).
2. Yapp, C. J. & Poeths, H. *Nature* **355**, 342–347 (1992).
3. Berner, R. A. *Science* **249**, 1382–1386 (1990).
4. Berner, R. A. *Am. J. Sci.* **291**, 339–376 (1991).
5. Berner, R. A. & Canfield, D. E. *Am. J. Sci.* **289**, 333–361 (1989).
6. Cerling, T. E. *Am. J. Sci.* **291**, 377–400 (1991).
7. Mora, C. I., Driese, S. G. & Seager, P. G. *Geology* **19**, 1017–1020 (1991).
8. Arthur, M. A., Allard, D. & Hinga, K. R. *Geol. Soc. Am. Ann. Meeting Abstr.*, A-178 (1991).
9. Freeman, K. H. & Hayes, J. M. *Global Biogeochem. Cycles* (in the press).
10. Hollander, D. J. & McKenzie, J. A. *Geology* **19**, 929–932 (1991).
11. Rau, G. H., Takahashi, T. & DesMarais, D. J. *Nature* **341**, 516–518 (1989).
12. Jasper, J. P. & Hayes, J. M. *Nature* **347**, 462–464 (1990).

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. They need not arise out of anything published in *Nature*. In any case, priority will be given to letters of fewer than 500 words and five references.

No easy despair

George Klein

The Fragile Species. By Lewis Thomas. *Scribners: 1992. Pp. 192. \$20.*

IN one of his earlier books, Lewis Thomas muses over the question of how our culture might introduce itself if we were to get a sudden chance to make ourselves known to a distant extraterrestrial civilization. Suppose we could send only a short message and that an answer could reach us only after several centuries. What would we send? But, of course. Some Johann Sebastian Bach. And then, as an afterthought: no, that is not possible. It would be bragging.

Sending Bach would certainly be bragging, and I am not sure it would be appreciated. I would rather send an essay by Lewis Thomas. It would be hard to find a more lucid introduction to the human condition. *The Fragile Species*, the fifth collection of the author's essays, continues in the tradition of his earlier books. It rolls forward gently and effortlessly like a mountain creek. Thomas is, by turns, deadly serious and gently teasing; sometimes ironic but never sarcastic; full of insight but never preaching; fuelled by deep emotion but never sentimental; precise and analytical yet a masterly synthesizer; intermittently reductionistic and holistic (sometimes in the same paragraph); equally fluent in science, medicine, literature, poetry and music; and able to empathize with every manifestation of animal and human life except one — boredom. I don't believe he knows what the word means.

The latest essays are once again the notes of a biology watcher. But Thomas's tone is less detached than in his previous books; the clouds are darker, the threats appear more ominous. AIDS and drug abuse are among the hard facts of life, and he does not refrain from calling a spade a spade. He abhors euphemisms such as 'substance abuse' or 'correctional institution'. I would add 'recreational drug' to his list. The increasingly younger ages of drug addicts, particularly in the poorest sections of the population, indicate that something has gone, in Thomas's words, "so badly wrong with the way we live together and the way children are treated in our society, that we should begin thinking about education in an entirely new light".

Thomas points out that our species is "brainy enough to menace all nature unless distracted by music". He reassures the weak-hearted, however, that our species is young and fragile, like an over-confident adolescent who does not

quite know what he or she is doing: "Growing up is hard times for an individual but sustained torment for a whole species, especially one as brainy and nervous as ours". We are new at the game and have not yet learned it. After we have shaken off this century, all might be running well again. Unjustified optimism? Perhaps, but Thomas's warning is clear enough for anyone willing to listen.

He also warns against triumphalism in medicine. As he reveals in the wonderful essays on "Becoming a Doctor" and "In Time of Plague", medicine is the eternal profession. Medical knowledge and technical savvy are biodegradable — since time immemorial, doctors have been trained to *do* things. For many centuries, the only hope for patients was to avoid — even by turning to homeopathy — unnecessary treatment by the medical profession, which adhered largely to the ancient idea of Galen that most diseases were usually due to congestion.

Now we live longer and more healthily, although we must concede that our general health has improved not because of medical science, but because of plumbbers and engineers. Our mental balance, however, has hardly improved; we are more afraid to die and less easily comforted. The steady increase in medical technology is inversely related to the

time doctors spend with each patient. Most of the time they seem to be somewhere else, worried about income, grants, tenure or parking.

But our lives are full of ameliorating circumstances. The sun is always breaking through the clouds in Thomas's sky. He remembers the abandoned term 'pleasure centre', identified accidentally by James Olds in the 1970s. Olds noticed that when rats were able to deliver an electric stimulus to a specific area in their brain, they would continue to do so for hours on end, forgoing food, drink and rest until they died of exhaustion. No matter where an enormous amount of reductionist research has taken the phenomenon, Thomas believes that the pleasure centre exists and still regards it as fair game for speculation. He does not mind that the problem became stuck, seeing it instead as "one of the loveliest of all the possibilities raised by biological science in memory". What was originally a side observation caused by the misplacement of an electrode has taught us that there is such a thing as pure pleasure. Jumping from science straight into the world of poetry as only he can do, Thomas suggests that the pleasure centre transmits to us and other vertebrates the great news that we are alive. And who knows, perhaps invertebrates have it too. If so, Thomas would "organize brass bands everywhere for dancing in all the streets of the world", until he ran out of money.

Contradicting the *Zeitgeist*, Thomas fearlessly states that major truths are often more easily pronounced by poets than by academic scientists. But one would be quite wrong to think that he has distanced himself from science. The



The formation of streams on slow-moving or stagnant ice in northern East Greenland. This picture comes from *Glaciers*, an illustrated study of the life and history of glaciers, by Michael Hambrey and Jürg Alean. Published by Cambridge University Press, price £19.95, \$29.95.

admiration he feels for the antibiotic revolution and the ever-increasing anchorage of clinical treatment in fundamental research is amplified when he looks at our era. Previously unapproachable problems such as cancer and the workings of the brain and the immune system have now turned into exciting puzzles where "cascades of surprise" are continuously pouring in from laboratories around the world. He reminds us that there has in fact been astonishing progress in the work on AIDS, especially considering that the human immunodeficiency virus has been known for less than a decade and is so sophisticated. Only "research, good, old-fashioned, obsessive, reductionistic science" can rid us of the disaster of AIDS, he says.

What is most important for Thomas among the innumerable topics that he treats so delightfully? I would single out symbiosis and communication. For him, there is nothing more comforting than the termite. Its nest is a model of perfect cooperation, and each insect is the equivalent of an eminently successful committee. Motile protozoans break down cellulose in its gut; spirochaetes attached to the protozoans' surface serve as their locomotory cilia; and each protozoan nurtures many symbiotic bacteria providing enzymes needed for digesting the wood eaten by the termite. This and innumerable other examples illustrate how potential adversaries, who once gave tit for tat, have become devoted collaborators. Soldiers entrenched on both sides of a front line during protracted war have shown similar tendencies. The cleft between struggle and cooperation is not as deep as one might think — altruism is just a special case among the various symbiotic arrangements in biology.

Human communication and language are two of Thomas's other recurrent preoccupations. He notes that the word 'pupil' stands for 'youngster' and 'pupil of the eye', not only in English, but also in Lapp, Chinese, Swahili and Samoan. His explanation for this universality rests on the fact that children can pick up any language with its proper pronunciation with remarkable ease. Looking into the eye of an attentive adult, a child would see his or her own reflected image and create the connection. This ability to acquire new language and make these kinds of connections allows Thomas to speculate, in parting, that language had its true origin in children.

It is because of such thought-provoking gems that I envy all those who have not yet read the book. They have a great pleasure in store. □

George Klein is in the Department of Tumour Biology, Karolinska Institute, Box 60400, 10401 Stockholm, Sweden.



Ascidians are abundant and highly diversified marine animals, and are our closest invertebrate relatives. This example of a 30-cm *Polycarpa clavata* is from Coral Reef Ascidians of New Caledonia. Published in English in France by Orstom, price FF300.

Friends and foes

Diana Barkan

Nationalism and Internationalism in Science, 1880–1939: Four Studies of the Nobel Population. By Elisabeth Crawford. Cambridge University Press: 1992. Pp. 157. £27.95, \$44.95.

THE 1970s witnessed a decline in traditional forms of international movement among scientists. Although exchanges increased among Western European countries, particularly among those of the European Communities, transatlantic interactions deteriorated. Many qualified scientists and engineers were meanwhile migrating from less developed countries to the West. By contrast, during the late 1980s there seems to have been a steady growth in transnational research ventures, investment and diversification; an increasing replacement of public with private funding; and a move towards local and regional research in science and technology centres in Europe and the United States. These trends are closely related to ideas about nationalism and internationalism in science, ideas which themselves are products of a long historical process. In this engaging book, Elisabeth Crawford explores the 'rise and fall' of internationalism in science between 1880 and 1939, addressing the complex relationship between national scientific élites and their governments, institutions and

foreign colleagues.

Crawford's case studies rely on data about the Nobel population, which between 1901 and 1939 numbered nearly 1,000 Nobel prize nominators and nominees in 25 countries. The workings of the Nobel institution rested on an international community of scientists who forwarded candidates for the prizes, so quantitative analysis of this population can provide a valuable indicator of trends in nationalism and internationalism in science.

Crawford shows that between 1880 and 1914 "nationalism and internationalism [in science] emerge both as friends and foes". She stresses the important distinction between "science as an abstract method for establishing universally valid knowledge" and "science as an activity and a social institution". As the former, science has promoted reproducibility and transmissibility across national boundaries. This "ethos of universalism" has often been mistakenly considered to be a supreme ruler in "all aspects of science . . . as a method, social activity, and institution, at all times." But, as her analyses show, during this century science has been primarily a national enterprise, used by the nascent nation-states to bolster economic, industrial and technological productivity as well as international standing. So a considerable gap has persisted between the practices and culture of national science and the rhetoric of universalism and internationalism.

Despite an increase in the number of
NATURE • VOL 358 • 9 JULY 1992

international congresses and meetings, an opening up of communication facilities and the establishment of the Nobel Prize in 1901, the trend to internationalism suffered severe retrenchment as a result of the First World War. This is the subject of Crawford's first case study. She shows how, in the "golden years of internationalism" between 1900 and 1914, Germans tended to be the most 'international' in their nominations of candidates for Nobel prizes; yet the years after the war saw an upsurge in nominations of own-country candidates by all scientists. And German scientists who had supported and participated in the war effort — in particular the much detested chemical weapons development — were barred for many years from international meetings. It was only after 1926 and the treaty of Locarno that the governmental, organizational and personal levels of scientific activity returned to normal in international relations.

In examining the relations between a scientific centre and periphery in Central Europe, Crawford challenges some sociological research that she finds lacking in historical sensitivity and specificity. Joseph Ben-David had argued that successful scientific development results from organizations built on competition, and that scientific groups removed geographically from a prominent centre invariably model themselves on that centre. But as Crawford indicates, Czech, Hungarian and Austrian groups of scientists developed research schools and even subdisciplines in their own countries independently of their colleagues in Germany — the scientific Mecca of the period.

The immediate and spectacular prestige of the Nobel prizes influenced the structure and aims of newly established research foundations, such as the Kaiser-Wilhelm Gesellschaft (now the Max-Planck Society). From its inception in 1911, the society created 'space' for prizewinners as administrators and advisers; the first directors of the institutes, A. Einstein, F. Haber, M. von Laue and R. Willstätter, had either already received a Nobel prize or were about to do so. As Crawford demonstrates, the society was able to further national science by building institutes for Germany's well-known scientists and fostering a climate of 'inbreeding' for nominations. The presence of prizewinners seems also to have alleviated the tension between basic and applied research that characterized the Kaiser-Wilhelm Gesellschaft when it was funded by the wealthy industrialists who became increasingly impatient after the war at the slower pace of applied work in the institutes.

Applied and industrial scientists were rarely rewarded with Nobel prizes,

although 25 per cent of US candidates in physics and chemistry were employed outside academic institutions. In her final case study, Crawford looks at the 'élite conception' of science in the United States and its role in the success of the national scientific enterprise. She suggests that a reevaluation is needed of H. Zuckerman's 1977 model of a sharply hierarchical, pyramidal US scientific community that forms institutions on the basis of accumulated advantage. In that model, US Nobel prizewinners made up an ultra élite, delineated from the rest of the academic élite (that is, even from the 1,000 or so scientists elected to the US National Academy of Sciences). Using archival data unavailable to Zuckerman, she finds that the similarities in educational background, master-apprentice relationship and employment history are much greater than any differences between winning and nonwinning Nobel candidates from 1901 to 1939.

In none of her studies does Crawford shy from considering developments in scientific theory and experiment, the peculiarities of the Swedish scientific community and the goals of the Nobel prize committees. Such a historical understanding must supplement any quantitative analysis of sociological variables if we are to appreciate how the prizes reflect national and international scientific relations. It is for this reason that her book is such a success. □

Diana Barkan is in the Division of the Humanities and Social Sciences, California Institute of Technology, Pasadena, California 91125, USA.

Film review

Andrew Holmes-Siedle

Hydrogenated Amorphous Silicon. By R. A. Street. Cambridge University Press: 1991. Pp. 417. £65, \$110.

IN 1976, the preparation of the first amorphous semiconductor diode was announced by W. Spear and P. LeComber of the University of Dundee. It was previously thought that one could not alter the conductivity of amorphous material by the addition of impurities (that is, by 'doping'). This meant that there was no way of making p-n junctions, the basis of nearly all semiconductor devices, from this material. Spear and LeComber found that if amorphous silicon films were prepared in a certain way, p-n junctions could be formed. I commented in *Nature* (260, 667 & 263, 458; 1976) that this finding seemed to be a breakthrough. In the 15 years since then, the enormous value of the discov-

ery has been proved, and research has confirmed that the films are not pure silicon but are better described as hydrogenated amorphous silicon.

The fact that thin films of the materials can be deposited over large areas means that the constraints of the crystal-growth method of making most semiconductors are avoided. This method normally gives wafers no larger than a few inches on which to make devices. Now that it is possible to dope and alter the structure during rapid film growth, the system has become a playground for the inventor of new devices. The faster the growth, the lower the cost of fabrication. Despite much research, the efficiency of doping as it affects conductivity is still much lower than for crystalline silicon. Even so, the material has assumed a firm place in the electronics industry.

Street and others at Xerox Palo Alto Research Center have developed theories for the mechanisms that give conductivity in disordered networks of atoms. Crystalline silicon conducts well because its atoms lie in an array of potential wells that are all of the same depth. This allows the electron wave function to be delocalized; that is, any electron has the run of the whole crystal. In irregular networks of atoms, there is a spread in the depth of the wells, causing electrons to be partly or completely localized. Preparation of films by glow-discharge methods permits some delocalization. Other influences also deter rapid movement of electrons in these materials. Disorder leads to collisions, and changes in the density of intrinsic defects, sometimes called 'dangling bonds', can temporarily trap electrons. The low conductivity of the first films produced seemed to confirm expectations that they might never give useful levels of conduction.

Street spends several chapters attempting to give the essence of the original breakthrough, focusing on the details of doping and conduction in hydrogenated silicon and germanium, in particular the co-ordination of the silicon and impurity atoms. Although I was keen to learn the tricks of the trade, I did not find it easy to draw the essence from the author's explanation. Like the atoms in question, the explanations are beautifully ordered in the short range but the long-range view is a bit obscure. For example, the curve of resistivity of crystalline silicon against dopant concentration appears nowhere in the book; and atom displacement by high-energy particles receives only a passing mention. This also applies to some items of history. Although the pioneering role of the Dundee group is acknowledged, the papers announcing many details of the group's findings (such as *Appl. Phys. Lett.* 28, 105; 1976 & *Phil. Mag.* 33, 935-9; 1976) are not

quoted, despite the inclusion of a graph of conductivity against doping that accompanied the group's initial brief announcement in 1975. And although a group at RCA, Princeton, was the first to announce the fabrication of p-i-n solar cells, useful photovoltaic responses in p-n junctions had been reported previously by workers at Dundee.

Street presents the subject as solid-state physics, but organic chemists might well feel at home with the book. His account would suggest that the amorphous silicon network has similarities to large organic molecules. Some of the same flexibility and individuality is present, implying that synthetic chemistry in thin silicon films may have as big a future as carbon chemistry has a past. The system is fertile in possibilities for new devices. Industry has responded: amorphous silicon solar cells — not the most efficient but adequate for domestic uses such as calculators and roof panels — are now mass-produced. Flat televi-

sion screens are operated by arrays of amorphous field-effect transistors. In the final chapter, Street reviews these applications and makes a few comments about the future. There is, however, little cross-referencing between this section and the rest of the book. It is also a surprise to find that the name of Street's colleague J. Mort, who coined the term "macroelectronics" for big amorphous device arrays, does not appear anywhere in the book.

This is a sober and well-organized review of the physical investigations that have been carried out into tetrahedral amorphous semiconductors. Although it is not an inspiring guide to the future of thin-film electronic devices, it is highly recommended to people entering the field professionally and to those who need an overview of the research so far.

Andrew Holmes-Siedle is the owner of REM, 64A Acre End Street, Eynsham, Oxford OX8 1PD, UK.

Illusions of empty space?

F. D. Kahn

The Fullness of Space: Nebulae, Stardust, and the Interstellar Medium. By Gareth Wynn-Williams. Cambridge University Press: 1992. Pp. 202. £35, \$65 (hbk); £15.95, \$29.95 (pbk).

Dust in the Galactic Environment. By D. C. B. Whittet. IOP/Hilger: 1992. Pp. 295. \$95, £47.50 (hbk); \$39, £22.50 (pbk).

To what extent should a book's title reflect its contents? Gareth Wynn-Williams has picked the phrase "The Fullness of Space" and no doubt expects us to recognize various allusions. His subtitle runs "nebulae, stardust and the interstellar medium", and the cover shows a picture of the Orion region, taken in the infrared part of the spectrum. One certainly gets the impression that this part of space is very crowded. The same region, more or less, reappears in two pictures on page 90, one taken at optical wavelengths, the other at 2.6-millimetre wavelength in radiation from molecules of carbon monoxide. These pictures give quite a different impression: indeed the optical photograph shows a scattering of stars and a handful of nebulae, and conveys the feeling of empty space — even though Orion is well populated by galactic standards. Pictures clearly need to be treated with great caution: what they reveal has been carefully processed and depends very much on the part of the spectrum used and on the way that the data are handled.

Douglas Whittet has chosen a title that is more explicit. His book also carries a striking picture on its cover, this time a view at optical wavelengths of the galaxy NGC891 riven apparently by a dark lane, and a set of contours is superposed to show the infrared emission from the same galaxy. With this image, the author neatly illustrates a basic concept. The dark rift is an illusion: in truth the dust in NGC891 extinguishes the light coming from the stars behind it. The energy thus gained is re-radiated by the dust in the infrared, so that this piece of sky is dark in the visible part of the spectrum but bright at longer wavelengths. The same picture illustrates another important concept: the shapes shown depend strongly on the resolution of the instrument used. The picture appears again in the first chapter of Whittet's book, this time with a circle giving the effective aperture of the satellite IRAS with which it was taken, and one sees why the infrared contours are so much wider spread than the rift in the optical picture.

It is easy to gain the impression from the image of NGC891 that the dust is hemmed in on all sides by the stars that surround it. To some extent this is true: the grain temperature is determined by the radiation field, and grains are destroyed in shock-waves, which are driven by stellar winds or supernova explosions. But the relationship is more complex. In reading Whittet's book, it soon becomes clear that its title could equally well have been "The Galactic Environment Determined by Dust". In general, dust is well mixed with the interstellar gas. When a cloud grows large enough, or becomes

sufficiently condensed, then its dust content also makes it opaque at visual wavelengths. Observation shows that such clouds contain hydrogen in molecular (H_2) rather than atomic (H) form. The dust acts as a catalyst in the formation of H_2 ; once this is accomplished, a laboratory full of other molecules can form as well. Here is a result that is not only interesting for its own sake but also important, because the molecular clouds are the very 'environment' in which new stars can form.

In a flight of fancy, Wynn-Williams asks what life would be like for us if we lived inside a molecular cloud. Our view of the Universe, at visual wavelengths, would end at a distance of about 0.25 light years. A sphere with this radius would contain just about enough diffuse matter to form another star of solar mass. So, with luck, we might expect to have another visible companion, in addition, of course, to the planets. It is a curious conclusion, but entirely justified, that in a space that was better filled we would have fewer neighbours that we were aware of. A rush-hour traveller on the London underground would probably understand this paradox.

By everyday standards, though, all interstellar densities are extremely low. Even a 'dense' cloud contains only about 10^{-20} g per cm^3 or 10^{-17} kg per m^3 (Whittet uses SI units, Wynn-Williams CGS units). What is it that changes at low densities and makes interstellar physics so different? The first change is that everything takes much longer: in a typical molecular cloud there would be an interval of about one year between successive collisions with a hydrogen molecule, and of course much longer for less abundant species. It takes a long time to achieve equilibrium between the different chemical species, or between the solid and the gaseous constituents. If at some stage it looks as though equilibrium had been attained, then along comes a shock to disturb the medium and move the goal-posts once again. Truly, the interstellar medium is a richly stocked lake in which to fish for problems.

Whittet's book is eminently suitable for those with serious intentions of learning about these matters and going on to do research at the professional level. Wynn-Williams' book, by contrast, is ideal for the interested layman and will fit nicely on a coffee table of modest size. Its text makes allowances for the nonspecialist, and even if the reader loses the thread of the argument, he or she can always enjoy the illustrations.

F. D. Kahn is in the Department of Astronomy, University of Manchester, Manchester M13 9PL, UK.

Modelling the hydrological cycle in assessments of climate change

D. Rind, C. Rosenzweig & R. Goldberg

Climate change caused by increasing atmospheric concentrations of greenhouse gases may have important effects on water circulation and availability and thus on agriculture, forestry and river flow, with significant economic consequences. A variety of models are being used to evaluate hydrological effects, but their hydrological responses to global warming are often inconsistent. Improved understanding of basic hydrological processes is needed if we are to assess the impact of future climate change.

THE current political debate concerning the necessary steps to limit greenhouse gas emissions depends crucially on the possible consequences of taking no action. Two questions arise: what will be the resulting climate change, and what will be the impact of that change. To address these questions, two Intergovernmental Panel on Climate Change (IPCC) working groups were established. The first (Working Group I) was assigned the task of identifying changes in climate resulting from increased concentrations of greenhouse gases. The second (Working Group II) was required to assess the environmental and socioeconomic consequences of climate change. Reports were issued by both groups in 1990^{1,2}, updates have recently been released, and it is intended that efforts should continue at least until the next century.

The hydrological cycle is central to both groups' concerns. Water vapour and cloud feedbacks currently provide the predominant positive amplifications of the direct CO₂-induced warming in general circulation models (GCMs)¹, increasing the direct CO₂ warming by 2–3 times. Changes in water availability directly affect agriculture, forestry, river flow and many different human activities, with important economic consequences. Clearly, our ability to determine the impacts of current greenhouse gas emissions depends heavily on our understanding of the hydrological cycle.

Models are important both for projecting the likely climate change and for assessing the environmental and socioeconomic response. Studies such as those reviewed by IPCC Working Group II and organized by the United States Environmental Protection Agency³ (EPA) have used a combination of climate and impact models. Climate change estimates from GCMs (generally temperature and precipitation) are used as the input to physical process models to produce estimates for quantities such as agricultural yields, forest biomass or river flow, and the resulting output is often used to evaluate the financial consequence. The predictions of these studies are alarming, with crop production estimated to decrease by some 30% and forest biomass by 40% in the United States³. But the results depend crucially on the hydrological cycles embedded in the different models; the issue of confounded hydrology is the focus of this review. We show that uncertainties in hydrological processes and inconsistencies in both climate and impact models limit our confidence in current assessments, and we suggest a future course of cooperative action.

A comparative study

For this review, we have chosen representative versions of several different types of models, all of which have contributed to climate impact assessments^{2–4}. For climate change we use the Goddard Institute for Space Studies (GISS) GCM⁵, for agricultural impact the CERES maize and wheat models^{6,7}, for forest response the FORENA forest dynamics model⁸ and for water

availability the Palmer drought severity index (PDSI) hydrological model⁹. Hydrologies were evaluated for both the current climate and a 4.5 °C temperature increase, for 15 different sites in the United States (Fig. 1). This temperature change is at the upper end of the projection for warming caused by CO₂ doubling, and clearly illustrates the differences between models. In the following tables results are shown from all 15 sites for the GCM, maize and PDSI models. Both the wheat model and forestry model could function at only eight sites with the prescribed warming; at the other seven, nothing could grow. Nevertheless, a comparison using the eight sites for which all models ran showed the overall characteristics reported here. Note that the GCM produces all elements of its hydrological cycle interactively, whereas the impact models are forced with specified precipitation and temperature (observed values are used for the current climate).

Although each of the models purports to represent the hydrological cycle with physically based parameterizations, the ground hydrological schemes that are used differ considerably from one another. Table 1 gives a summary of the various parameterizations and timescales involved. Differences arise in every category; for example, different formulae are used to calculate potential evapotranspiration (PET), which is the maximum evaporation possible under given atmospheric conditions. The actual evaporation may or may not be a function of ground wetness or leaf area index, depending on the model. In some cases runoff is calculated as a direct function of precipitation, whereas in others it is calculated only when the soil is completely saturated. Given such differences, it would be no surprise if the models' responses to climate change varied considerably.

The annual average hydrological cycles for the current climate from the models are presented in Table 2a, together with observations from various sources^{10–14}. Averaging the cycles annually and over diverse geographical regions removes some of the discrepancies between the models; it also allows us to illustrate succinctly the basic conclusions, which are not altered by this procedure. The observations of precipitation and runoff are relatively well constrained by measuring gauges; evaporation is a calculated quantity, which can be checked to ensure hydrological balance with precipitation and runoff (for the long-term annual average this is probably realistic). Soil moisture values are the result of water budget analyses¹³; although this procedure can produce realistic results when compared with *in situ* observations¹⁵, it depends on the accuracy of all the other observations and the PET formulation.

PET is a hypothetical variable, indicative of the maximum amount of evaporation possible in a given circumstance. In most formulations it can be constrained by pan evaporation measurements¹⁴ over a completely wet surface. But the aerodynamic formulations that are used in GCMs calculate PET above

TABLE 1 Hydrological processes in climate and impact models

	Parameter	GISS GCM	CERES Wheat, maize	FORENA	Palmer drought severity index
Precipitation Method Time interval	P	Calculated Hour	Input Daily observed	Input Long-term monthly observed	Input Monthly observed
Runoff Infiltration	R	$f(\text{SM}, P)$ $P - R - E$	$f(\text{Soil type, SM}, P)$ $P - R$	$\text{SM} > \text{soil water capacity}$ P until $\text{SM} > \text{soil water capacity}$	$\text{SM} > \text{soil water capacity}$ P until $\text{SM} > \text{soil water capacity}$
Potential evaporation	PET	Aerodynamic ⁵	Priestley-Taylor ²⁷	Thorndthwaite ²⁸	Thorndthwaite ²⁸
Total evaporation	E	$f(\text{PET, ground wetness})$	$f(\text{PET, LAI, time})$	$f(\text{PET, ground wetness})$	PET ($\text{SM} > 0$)
Drainage	D	none	$\text{SM} > \text{soil water capacity, bottom layer}$	none	none
Soil moisture	SM	$P - E - R$	$P - E - R - D$	$P - E - R$	$P - E - R$
Area of calculation		$10^5\text{--}10^6 \text{ km}^2$	Single plant*	$10^2\text{--}10^3 \text{ m}^2$ *	$10^4\text{--}10^5 \text{ km}^2$ †

All parameters used in the table are given in column 2, except LAI (leaf area index).

* Assumes uniform stand.

† By state, as used here.

partially wet surfaces over large geographical areas and produce significantly higher values (Table 2a), as the partially wet surfaces have higher temperatures than water in a pan.

The current climate values are similar for many quantities. The GCM has perhaps too high a ratio of evaporation to runoff, and the FORENA model has too much soil moisture (which is strongly dependent on the soil type used), but overall the basic hydrological variables are reasonably close to the observed values (which are only approximate in any case). It is likely that all the models are calibrated to the current climate in their development mode, given the observed precipitation which thus must be balanced by evaporation, runoff and soil moisture storage. Not surprisingly, the greatest discrepancies are found in the soil moisture (which is the least well-observed quantity) and potential evapotranspiration (which is calculated differently in the different models); the larger value listed for the GCM is the aerodynamic PET used to calculate the model's evaporation⁴, whereas the smaller number is a diagnostic calculation of the Penman PET¹⁶.

We now consider how the different models respond to a 4.5 °C warming. The impact models were re-run with temperatures warmed uniformly by this amount, and no change in precipitation from current climate values (to simplify the analysis); the GCM changes are produced by warming the sea surface temperatures by 4.5 °C everywhere, which results in little change in the annual average precipitation. Note that in contrast with the customary procedure, we are not using the GCM-generated changes in the impact models but are forcing all the models with a specified warming to illustrate their hydrological sensitivity.

The annual results are shown in Table 2b, with the number

in parentheses representing the percentage change from the control run. Large disparities are evident: the PDSI indicates a 35% decrease in soil moisture, whereas the maize model shows a slight increase; the PDSI predicts a 46% decrease in runoff, and the forestry model a 24% decrease, in contrast to the much smaller changes in the agricultural models (owing to their lack of soil moisture change). The GCM results, produced by warming the sea surface temperatures by 4.5 °C everywhere, are often intermediate in magnitude.

Although these models do not all have hydrological cycle changes as their primary output, such changes affect the key results that the models were designed to produce. For example, if the maize model were to experience the runoff and soil moisture deficits associated with the PDSI or forestry models, it would probably show greater reductions in maize yields. As the PDSI was designed for agricultural assessments¹⁷, such discrepancies are not to be taken lightly.

The differences in response are associated to a high degree with the governing parameterizations or assumptions embedded in each model (Table 1). For example, the Thorndthwaite PET incurs a higher percentage increase than does the Priestley-Taylor indicating that a greater percentage increase in evaporation is possible (and does in fact occur) with the Thorndthwaite formulation. The PDSI and the GCM have very different evaporation efficiencies, with the PDSI converting increased atmospheric demand to evaporation at ~90% efficiency, whereas the GCM has ~20% efficiency, owing primarily to differences in the empirical and aerodynamic PET formulations⁴.

The models also contain very different assumptions. The agricultural models have extremely low values of soil moisture loss except during their explicit growing season. Thus, during



FIG. 1 Stations used in the comparison of model output.

TABLE 2 Annual hydrological cycle and change predicted for 4.5 °C warming

(a) Annual hydrological cycle for the current climate						
Source	Precipitation (mm per day)	Evaporation (mm per day)	Runoff (mm per day)	Soil moisture (mm)	Potential evaporation (mm per day)	
Observations	2.30	1.55	0.70	78.2	3.16	
GCM	2.55	2.07	0.49	98.0	10.4/21.7	
PDSI	2.30	1.67	0.64	113.0	2.32	
Maize	2.30	1.56	0.77	130.0	3.77	
Wheat	2.69	1.80	0.91	151.3	3.85	
Forest	2.92	2.13	0.89	293.6	2.28	
(b) Change in annual hydrologic cycle with a 4.5 °C warming						
Source	Temperature (°C)	Precipitation (mm per day)	Evaporation (mm per day)	Runoff (mm per day)	Soil moisture (mm)	Potential evaporation (mm per day)
GCM	4.43	0.003	−0.01 (1%)	−0.07 (14%)	−4.1 (4%)	2 (17%)/10 (48%)
PDSI	4.50	0	0.29 (18%)	−0.30 (46%)	−39.9 (35%)	0.86 (37%)
Maize	4.50	0	0.02 (2%)	−0.03 (3%)	0.7 (0.5%)	0.61 (16%)
Wheat	4.50	0	0.02 (2%)	−0.08 (9%)	−3.3 (2%)	0.94 (24%)
Forest	4.50	0	0.27 (13%)	−0.21 (24%)	−32.3 (11%)	0.89 (39%)

Results from the wheat and forest models come from eight sites; all other results are from the full 15 sites.

July when temperatures are warmest, the 4.5 °C increase provides little additional evaporation in the wheat model despite a soil moisture increase, in contrast to the response in the other models (Fig. 2). This is because wheat is not growing, and roots are not present to extract water from lower soil layers. These results imply that if forests and wheat fields were to be located in the same region, large horizontal gradients in soil moisture would exist under equilibrium conditions. The forestry model assumes that at the beginning of the year (January) soil moisture is returned to saturation conditions both in the current climate and in changed climates, a hypothesis which is far from certain with large climate alterations. When the model was re-run without this assumption, larger soil moisture deficits occurred in some regions, with results closer to the PDSI values.

Future directions

This comparative study raises many questions and suggests much-needed improvements. Primarily, we must increase our understanding of the hydrological cycle; different formulations are possible in models because of the lack of general consensus on the physical processes and their parameterizations. In particular, we need to know how to model many of the processes depicted in Table 1. Can runoff and infiltration be calculated accurately from models that have relatively coarse vertical and horizontal resolution? Does small-scale heterogeneity overwhelm averaged values of surface parameters, used in these models by necessity, and how do we model sub-grid-scale soil moisture (or precipitation)? What is the proper formulation for PET? How accurate is the crop model's low estimate of soil moisture loss for high temperature in the non-growing season? Can large horizontal gradients in soil moisture exist between a forested and agricultural landscape for long periods of time, as

implied by these results (Fig. 2)? Is drainage through a permeable bottom layer necessary for accurate representation of surface hydrology (as shown in Table 1, it is included only in the crop models)?

Another possible source of hydrological inconsistencies arises from the different scale sizes used by the models. There are obvious difficulties associated with going from the large scale of the GCM grid boxes (hundreds of kilometres) to the small scales of the physical processes (hundreds of metres), or even down to the level of a single plant (Table 1). Various approaches are being tested, including the interpolation of grid box averages to the smaller scales, the use of empirical relationships between the large and small scales for the current climate, and the embedding of finer-resolution grids within the GCM at particular locations. Although all of these techniques are subject to error, they should be evaluated so as to provide a range of potential forcings.

Other translation problems exist. Because of the difficulties in modelling current precipitation using GCMs, the ratio of the GCM's new-climate to present-climate precipitation is often used to estimate precipitation change, but this approach does not allow for changes in precipitation frequency. It is hoped that improvements in aspects of GCMs such as their physical representation of the land surface¹⁸ and convection¹⁹ will ultimately eliminate the need for this.

The versions of these models that we have compared are those used in the EPA study on the impact of climate change in the United States^{20,21}. Most of the models were not originally intended for the purpose of climate-change assessment, and although new versions are being developed, it is necessary to guarantee that we establish a suite of models with consistent hydrological representations and assumptions if they are to be used together.

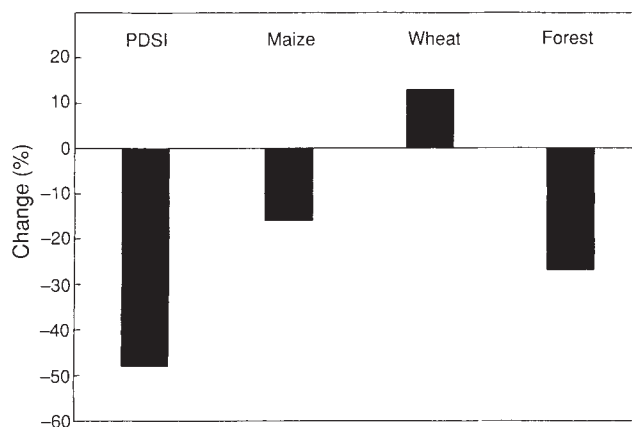


FIG. 2 Soil moisture changes during July caused by 4.5 °C temperature increase in the various impact models.

REVIEW ARTICLE

The temperature and precipitation forecasts of GCMs are strongly affected by the soil moisture changes, which are directly associated with aspects of their ground hydrology schemes (for example control-run values of soil moisture²², or run-off over frozen ground²³); can it really be sensible to use GCM predictions in models that then calculate a very different soil moisture response?

The need to understand hydrological processes better has motivated the Global Energy and Water Cycle Experiment (GEWEX)²⁴, a component of the World Climate Research Program, which has as an initial goal the observation and modelling of the Mississippi River basin beginning in 1993. In the United States, the Interagency Committee on Earth and Environmental Sciences (CEES) has identified the hydrological cycle in general as the highest scientific priority for global change research. Intensive field studies, of the nature of FIFE²⁵, HAPEX and others²⁶, will be necessary for a wide variety of terrains in order to answer these basic questions.

1. Intergovernmental Panel on Climate Change Working Group I *Climate Change, the IPCC Scientific Assessment* (Cambridge Univ. Press, Cambridge, 1990).
2. Intergovernmental Panel on Climate Change Working Group II *Climate Change, the IPCC Impacts Assessment* (Australian Government Publishing Service, Canberra, 1990).
3. Smith, J. B. & Tirpak, D. EPA-230-05-89-050 (U.S. Environmental Protection Agency, Washington DC, 1989).
4. Rind, D. *et al. J. geophys. Res.* **95**, 9983-10004 (1990).
5. Hansen, J. *et al. Mon. Weath. Rev.* **111**, 609-662 (1983).
6. Jones, C. & Kiriya, J. *CERES-Maize* (Texas A & M Press, College Station, 1986).
7. Ritchie, J. T. & Otter, S. in *ARS Wheat Yield Project*, ARS-38 (ed. Willis, W. O.) 159-175 (Department of Agriculture, Washington DC, 1985).
8. Shugart, H. H. *A Theory of Forest Dynamics* (Springer, New York, 1984).
9. Palmer, W. C. *U.S. Weather Bureau Res. Pap.* 45 (Library and Information Service Division, NOAA, Washington DC, 1965).
10. *National Water Summary Water Supply Pap.* 2350 (U.S. Geological Survey, U.S. Government Printing Office, 1990).
11. Rudolf, W. *World Climate* (Wissenschaftliche Verlagsgesellschaft, Stuttgart, 1981).
12. Korzun, V. I. *et al. (eds) Atlas of World Water Balance* (UNESCO Press, 1977).
13. Willmott, C. J., Rowe, C. M. & Mintz, Y. *J. Climatol.* **5**, 589-606 (1985).
14. Farnsworth, R. K. & Thompson, E. S. *NOAA Tech. Rep. NWS* 33, 34 (NOAA, Washington DC, 1982).
15. Robinson, J. M. & Hubbard, K. G. *Agron. J.* **82**, 1141-1148 (1990).
16. Penman, H. L. *Proc. R. Soc. A* **193**, 120-146 (1948).

GUIDE TO AUTHORS

PLEASE follow these guidelines so that your manuscript may be handled expeditiously.

Nature is an international journal covering all the sciences. Contributors should therefore bear in mind those readers who work in other fields and those for whom English is a second language, and write clearly and simply, avoiding unnecessary technical terminology. Space in the journal is limited, making competition for publication severe. Brevity is highly valued. One printed page of *Nature*, without interruptions of the text, contains about 1,300 words.

Manuscripts are selected for publication according to editorial assessment of their suitability and reports from independent referees. They can be sent to London or Washington and should be addressed to the Editor. Manuscripts may be dealt with in either office, depending on the subject matter, and with where necessary be sent between offices by overnight courier. All manuscripts are acknowledged on receipt but fewer than half are sent for review. Those that are not reviewed are returned as rapidly as possible so that they may be submitted elsewhere without delay. Contributors may suggest reviewers; limited requests for the exclusion of specific reviewers are usually heeded. Manuscripts are usually sent to two or three reviewers, who are chosen for their expertise rather than their geographical location. Manuscripts accepted for publication are typeset from the London office.

Nature requests authors to deposit sequence and crystallographic data in the databases that exist for this purpose, and to mention availability of these data.

Once a manuscript is accepted for publication, contributors will receive galley proofs in about 4 weeks. *Nature's* staff will edit manuscripts with a view to brevity and clarity, so contributors should check their proofs carefully. Manuscripts are generally published 2-3 weeks after receipt of corrected proofs. *Nature* does not accept page charges. Contributors receive a reprint order form with their proofs; reprint orders are processed after the manuscript is published and payment received.

Categories of paper

Review articles survey recent developments in a field. Most are commissioned, but suggestions are welcome in the form of a one-page synopsis addressed to the Reviews Coordinator. Length is negotiable in advance.

Articles are research reports whose conclusions are of general interest and which are sufficiently rounded to be a substantial advance in understanding. They should not have more than 3,000 words of text (not including figure legends) or more than six display items (figures and tables) and should not exceed more than five pages of *Nature*.

Articles start with a heading of 50-80 words written to advertise their content in general terms, to which editors will pay particular attention. The heading does not usually contain numbers, abbreviations or measurements. The introduction to the study is contained in the first two or three paragraphs of the article, which also briefly summarize its results and implications. Articles have fewer than 30 references and may contain a few subheadings of two or three words.

Letters are short reports of outstanding novel findings whose implications are general and important enough to be of interest to those outside the field. Letters should have 1,000 or fewer words of text and four or fewer display items. The first paragraph describes, in not more than 150 words and without the use of abbreviations, the background, rationale and chief conclusions of the study for the particular benefit of non-specialist readers. Letters do not have subheadings and contain fewer than 30 references.

Commentary articles deal with issues in, or arising from, research that are also of interest to readers outside research. Some are commissioned but suggestions can be made to the commentary editor in the form of a one-page synopsis. Commentaries are normally between one and four pages of *Nature*.

News and Views articles inform non-specialist readers about new scientific advances, sometimes in the form of a conference report. Most are commissioned but proposals can be made in advance to the

News and Views editor.

Scientific Correspondence is for discussion of topical scientific matters, including those published in *Nature*, and for miscellaneous contributions. Priority is given to letters of fewer than 500 words and 5 references.

Preparation of manuscripts

All manuscripts should be typed, double-spaced, on one side of the paper only. An original and three copies are required, each accompanied by artwork. If photographs are included, four sets of originals are required; for line drawings, one set of originals and three good-quality photocopies are acceptable. Reference lists, figure legends and tables should all be on separate sheets, all of which should be double-spaced and numbered. Relevant manuscripts in press or submitted for publication elsewhere should be included with each copy of a submitted manuscript, and clearly marked as such. Revised and resubmitted manuscripts should also be clearly marked as such and labelled with their manuscript numbers.

Titles say what the paper is about with the minimum of technical terminology and in fewer than 80 characters in the case of Articles and Letters. Active verbs, numerical values, abbreviations and punctuation are to be avoided. Titles should contain one or two key words for indexing purposes.

Abstracts should be marked individually and clearly with the author's name and, when known, the manuscript number. Ideally, no figure should be larger than 28 by 22 cm. Figures with several parts are to be avoided and are permitted only if the parts are closely related, either experimentally or logically. Unlettered original photographs should be provided. Original artwork is returned when a manuscript cannot be published.

Protein/nucleotide sequences should ideally be in the three-letter and not the single-letter code for amino acids. One column width of *Nature* can accommodate 30 amino acids or 60 base pairs. Numbering of sequences should be in the left-hand margin only, with a single space between rows.

Suggestions for cover illustrations, with captions and labelled with the manuscript number, are welcome. **Colour artwork.** A charge of £500 per page is made as a contribution towards the cost of reproducing colour figures. Inability to pay these costs will not prevent the publication of essential colour figures if the circumstances are explained. Proofs of colour artwork may be sent to contributors under separate cover from their galley proofs. Figure legends should not exceed 300 words and ideally should be shorter. The figure is described first, then, briefly, the method. Reference to a method published elsewhere is preferable to a full description. Methods are not described in the text.

References are numbered sequentially as they appear in the text, followed by those in tables and finally by those in figure legends. Only papers published or in the press are numbered and included in the reference list. All other forms of reference, including unreferenced abstracts, should be cited in the text as a personal communication, manuscript submitted or in preparation. Text is not included in reference lists. References are abbreviated according to the *World List of Scientific Periodicals* (Butterworths, London, 1983-85).

The first and last page numbers are included; reference to books should include publisher, place and date. **Abbreviations**, symbols, units and Greek letters should be identified the first time they are used. Acronyms should be avoided whenever possible and, if used, defined. Footnotes are not used except for changed addresses. **Acknowledgements** are brief, grant and contribution numbers are not allowed. **Submission.** Manuscripts can be sent to the Editor at 4 Little Essex Street, London WC2R 3LF, UK or at 1137 National Press Building, Washington, DC 20045, USA. Manuscripts or proofs sent by air courier to London should be declared as 'manuscripts' and 'value \$5' to prevent the imposition of import duty and value-added tax (VAT).

To engender consistent and appropriate models, we must foster a community of modellers, concerned both with climate changes and with impacts (IPCC Working Groups I and II). Early communication between these two groups has consisted of the development of 'wish lists' of GCM variables needed for impact studies. The results discussed here reveal some of the weaknesses of this approach, and emphasise that the next stage of communication must concern processes. Interactions are also necessary between modellers in the impact community; with models ostensibly operating for the same site, how much heterogeneity in response is plausible? The climate change and climate impact communities must together use whatever new understanding of hydrological processes arises from present or future programme if we are to improve our assessments of the impact of climate change. □

D. Rind is at the NASA/Goddard Institute for Space Studies, New York, New York 10025, USA. C. Rosenzweig and R. Goldberg are at the Center for Global Habitability, Columbia University, New York, New York 10027, USA.

17. NOAA National Weather Service *Handbook No. 1*. (NOAA Washington DC, 1979).
18. Abramopoulos, F., Rosenzweig, C. & Choudhury, B. *J. Clim.* **1**, 921-945 (1988).
19. Del Genio, T. & Yao, M.-S., in *Cumulus Parameterization* (eds Emanuel, K. A. & Raymond, D. A.) (American Meteorological Society, in the press).
20. Adams, R. *et al. Nature* **345**, 219-224 (1990).
21. Overpeck, J. & Bartlein, P. J. in *The Potential Effects of Global Climate Change on the United States, Appendix D. Forests* (eds Smith, J. & Tirpak, D.) 1-1-1-32 (U.S. Environmental Protection Agency, Washington DC, 1990).
22. Meehl, G. & Washington, W. *J. Atmos. Sci.* **45**, 1476-1492 (1988).
23. Mitchell, J. F. B. & Warrilow, D. A. *Nature* **330**, 238-240 (1987).
24. Chahine, M. T. *EOS* **73**, 9-14 (1992).
25. Sellers, P. J., Hall, F. G., Asrar, G., Strebel, D. E. & Murphy, R. E. *Bull. Am. met. Soc.* **69**, 22-27 (1988).
26. National Research Council, Working Group 5 in *Research Strategies for the U.S. Global Change Research Program* 131-163 (National Academy Press, Washington DC, 1990).
27. Priestley, C. H. B. & Taylor, R. *J. Mon. Weath. Rev.* **100**, 81-92 (1972).
28. Thornthwaite, C. W. *Geog. Rev.* **38**, 55-89 (1948).

ACKNOWLEDGEMENTS. We thank R. Webb for useful suggestions on an earlier version of the manuscript. This work was supported by the U.S. Environmental Protection Agency Global Climate Change Division, Office of Policy Analysis, and climate modelling at GISS is supported by the NASA Climate Program Office.

Copies of articles from this publication are now available from the UMI Article Clearinghouse.

For more information about the Clearinghouse, please fill out and mail back the coupon below.

UMI Article Clearinghouse

Yes! I would like to know more about UMI Article Clearinghouse.

I am interested in electronic ordering through the following system(s):

- ☐ DIALOG/Dialorder ☐ ITT Dialcom
☐ OnType ☐ OCLC ILL Subsystem

☐ Other (please specify) _____

☐ I am interested in sending my order by mail.

☐ Please send me your current catalog and user instructions for the system(s) I checked above.

Name _____

Title _____

Institution/Company _____

Department _____

Address _____

City _____ State _____ Zip _____

Phone (_____) _____

Mail to: University Microfilms International
300 North Zeeb Road, Box 91 Ann Arbor, MI 48106

RNA substrate binding site in the catalytic core of the *Tetrahymena* ribozyme

Anna Marie Pyle*, Felicia L. Murphy & Thomas R. Cech†

Howard Hughes Medical Institute and Department of Chemistry and Biochemistry, University of Colorado, Boulder, Colorado 80309, USA

In catalysis by group I introns, the helix (P1) containing the RNA cleavage site must be positioned next to the guanosine binding site. We have identified a conserved adenine in the catalytic core that contributes to the stability of this arrangement and propose that it accepts a hydrogen bond from a specific 2'-OH in P1. Such base-backbone tertiary interactions may be generally important to the organization of RNA tertiary structure.

SHORTENED forms of the *Tetrahymena thermophila* pre-ribosomal RNA intron can catalyse sequence-specific cleavage of oligonucleotides using a guanosine nucleophile¹⁻³. Cleavage is preceded by base-pairing of the oligonucleotide substrate to a complementary internal guide sequence (IGS) at the 5'-end of the RNA enzyme (ribozyme, Fig. 1)^{1,4-7}. The resulting helix (P1) is located far from the guanosine binding site in the secondary structure of the molecule⁸, so tertiary structure presumably holds these two elements together. The results described here indicate that P1 binding is mediated in part by interaction between a specific ribose 2'-OH on the substrate strand of P1 and a particular adenine in the phylogenetically conserved core. This interaction is especially interesting because it involves recognition of a ribose 2'-OH group within double-helical RNA, and thereby provides a mechanism for packing an RNA helix into a folded RNA structure.

Substrate-ribozyme tertiary interactions were first implicated by the finding that oligoribonucleotides bind to the ribozyme with an energy much greater than expected for base-pairing interactions alone⁹⁻¹³ and are consistent with the ability of a detached P1 helix to serve as a substrate for cleavage by the core of the ribozyme¹⁴. Specific interactions were localized to 2'-OH groups on the sugar residues 2 and 3 nucleotides from the substrate cleavage site (positions -2c and -3u, Fig. 1)^{15,16}. These independent interactions contribute 1 and 2 kcal mol⁻¹ of extra binding energy, respectively, at 42 °C. The magnitude of this energetic stabilization suggests that the 2'-OH groups form discrete bonds to the ribozyme that should be identifiable.

After inspection of the three-dimensional model of Kim and Cech¹⁷, taking into consideration the guanosine binding site and stabilizing base-triple interactions that had subsequently been identified^{8,18}, we chose specific bases to evaluate for their contribution to tertiary interactions with substrate. Bases implicated in the binding of the guanosine cofactor⁸ were disregarded as candidates for the binding of substrate, because G and oligoribonucleotide binding are largely independent^{12,13}. Conversely, a few phylogenetically conserved bases (at positions 301-303) exhibited behaviour consistent with involvement in 5'-exon binding¹⁹ (analogous to substrate binding) and were therefore possible points of tertiary interaction. The latter class of residues, and a number of other highly conserved residues in the core for which no function was known, were selected as candidates for tertiary contact to the RNA substrate.

Effect localized to A302

A gel-shift binding assay¹² was used to screen mutant ribozymes for their ability to distinguish between the all-RNA oligonucleotide MP (GGCCUCU) and oligonucleotides MPd3 and MPd2,

which have single deoxynucleotide residues three and two nucleotides from the cleavage site (GGCCCU(U)CU and GGCCCUd(C)U, respectively). The oligonucleotide product rather than substrate was used to preclude reaction, as RNA product and substrate are similar in their binding properties^{12,13}. The strategy for the screen was as follows. There is a 33-fold difference in the dissociation constant (K_d) of MP and MPd3 binding to the wild-type (WT) ribozyme (Table 1, column 3), representing a $\Delta\Delta G^\circ$ of 2.2 kcal mol⁻¹ (Fig. 2). Any mutant which showed equivalent binding of MP and MPd3, while still discriminating MP from MPd2, is specifically defective in recognition of the -3u 2'-OH group. The base normally present at the position of mutation would therefore be a strong candidate for tertiary contact to the -3u 2'-OH group on bound oligonucleotide. Analogous experiments could identify bases involved in tertiary contact to the -2c 2'-OH group.

The first sets of mutants generated and screened were A114N, A207N, A301N, G303N, and A306N, where A306N indicates that the A at position 306 has been changed to the three other bases (Table 1). Oligonucleotide binding by the A306N mutants was similar to that of the WT ribozyme. Although absolute levels of oligonucleotide binding varied for A114N and A207N mutants, they showed $\Delta\Delta G^\circ$ values for differentiation of MPd3

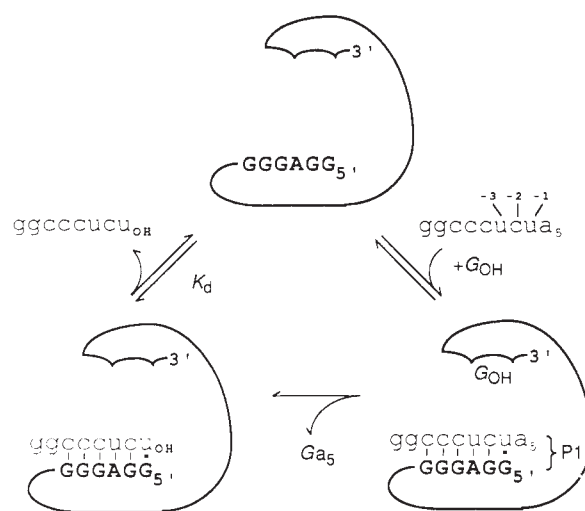


FIG. 1 Pathway for 5'-ggccucuaaaaa (matched substrate, or MS) recognition and reaction by the *Tetrahymena* ribozyme. Substrate nucleotides are numbered from the cleavage site as shown. The guanosine cofactor (G) enters its binding site and catalyzes cleavage of MS, with loss of Gaaaaa. Both the substrate and product oligonucleotides are in equilibrium with the free ribozyme. The K_d shown, for product oligonucleotide, is the quantity measured and described in Fig. 2 and Table 1.

* Present address: Department of Biochemistry and Biophysics, Columbia University, New York, New York 10032, USA.

† To whom correspondence should be addressed.

TABLE 1 Parameters used to evaluate the effects of ribozyme mutation on tertiary contacts

Mutant	K_d of MP (nM)	Ratio of K_d^* MPd3/MP	Ratio of K_d MPd2/MP	Ratio of K_d MPd3/MPd2	$(k_{cat}/K_m)^{S\dagger}$ ($M^{-1} min^{-1}$)	Ratio of $(k_{cat}/K_m)^S$ dMSr3/dMS
WT	0.3 $\ddagger$	33	6.3	5.3	6,600	13
A301U	17	11	7.6	1.4	100	8.2
A301C	26	8.8	3.4	2.6	88	5.0
A301G	33	3.5	2.9	1.2	650	1.5
A302U	2.1	7.1	12	0.59	800	1.9
A302C	56	2.5	3.7	0.67	29	—
A302G	53	1.2	2.9	0.41	1,400	0.45
G303U	18	6.2	1.8	3.4	22	13
G303C	64	5.3	2.3	2.3	—	—
G303A	5.2	16	5.6	2.9	240	13
A304U	1.0	61	10	6.1	810	20
A304C	0.69	30	4.6	6.5	950	24
A304G	1.6	39	8.1	4.8	520	20
A306U	2	14	6.2	2.2	—	—
A306C	2	21	7.7	2.7	—	—
A306G	2	11	3.6	3.1	—	—
A114U	38	12	2.4	5.2	83	4.5
A114C	64	12	2.0	6.0	33	17
A114G	8.0	30	9.1	3.3	45	21
A207U	18	11	5.2	2.2	—	>1.0
A207C	1.6	36	40	0.9	160	23
A207G	17	11	8.9	1.2	79	4.5

All binding constants shown have been obtained from binding curves of at least nine points and each curve has been repeated at least once, in some cases by K_{rel} comparison³⁷. A gel electrophoresis method was used to measure binding. Experimental conditions and error limits for values have been described previously^{12,15}. In these experiments, MP, MPd2 and MPd3 concentration = 0.2 nM, Mg^{2+} = 10 mM, equilibration time was 10–15 min (1–4 h in the case of WT), temperature = 42 °C. MP, Matched product GGCCCUU; MPd3, GGCCCU(U)U; MPd2, GGCCCUd(C)U. Values for the A306N mutants were obtained by comparison of MP, MPd2 and MPd3 binding at single ribozyme concentrations.

* Division of the K_d for MPd3 binding (data not shown) by the K_d for MP binding (shown in column at left) yields these values. These value are converted to $\Delta\Delta G^\circ$ and graphed in Fig. 2. As $K_d(MPd3)/K_d(MP)$ approaches 1 or $\Delta\Delta G^\circ$ approaches 0, the position of mutation is likely to be a position of tertiary contact with the substrate 2'-OH group three nucleotides from the cleavage site.

† The second order rate constant for cleavage of d(CCCUCUA₅) (dMS) by the ribozymes indicated at left. Conditions are ribozyme excess and G excess, with trace amounts of S, using a procedure similar to that published previously¹³. Six timepoints up to 4 h were taken under conditions of 0.25 μ M ribozyme (E), 3 mM G, 10 mM Mg^{2+} , 2 nM ³²P-dMS, 42 °C. The incubation buffer used was the same as for the binding studies: 50 mM Tris, 0.1 mM EDTA, 10 mM NaCl, 10 mM $MgCl_2$, pH 7.5. That this large screen was performed under $(k_{cat}/K_m)^S$ conditions (where the concentration of E is below saturation with S) was suggested by obtaining complete K_m curves for WT and A302G cleavage of dMS, dMSr3 and dMSf3 at seven different E concentrations (0.05 to 4 μ M).

‡ This number was previously observed to be 0.8 nM¹⁵. At longer equilibration times (1–4 h) and using MP concentration of 0.05 nM, much lower than the K_d , binding was observed to be somewhat tighter, as shown here. The K_d of d(GGCCUUCU) binding to the ribozyme at 42 °C is 2,100 nM, as reported previously.

from MP similar to that of the WT ribozyme (Table 1, Fig. 2). The A301N and G303N mutants showed some partial effects on $\Delta\Delta G^\circ$, suggesting that they might be indirectly involved in a stabilizing contact.

Mutagenesis was therefore done on two less conserved bases in the core that were proximal to A301 and G303. One of these, A304, is phylogenetically conserved at the 89% level²⁰. Mutations at this position had little effect on activity or substrate recognition. The other base, A302, is conserved except for a 20% variation to U. The A302G mutant differentiated only slightly between MPd3 and MP, with a $\Delta\Delta G^\circ$ of 0.1 kcal mol⁻¹ compared to 2.2 kcal mol⁻¹ for the WT ribozyme (Fig. 2). $\Delta\Delta G^\circ$

for A302C differentiation was 0.6 kcal mol⁻¹. Thus, A302G and A302C, unlike all the other mutants tested, had the properties expected if the mutagenized base formed a tertiary contact to the -3u 2'-OH group (Fig. 2). Differentiation by A302U ($\Delta\Delta G^\circ$ of 1.2 kcal mol⁻¹) was only partially affected. This suggests that the ribozyme with a uracil at position 302 can weakly interact with the -3u 2'-OH group. The observation that U can partially substitute for A is consistent with the phylogenetic data.

When the putative contact is removed by alteration of the ribozyme, the K_d increases from 0.3 nM (for WT binding to MP) to about 50 nM (for A302C or A302G binding to MP). This destabilization can be compared to the energy lost when

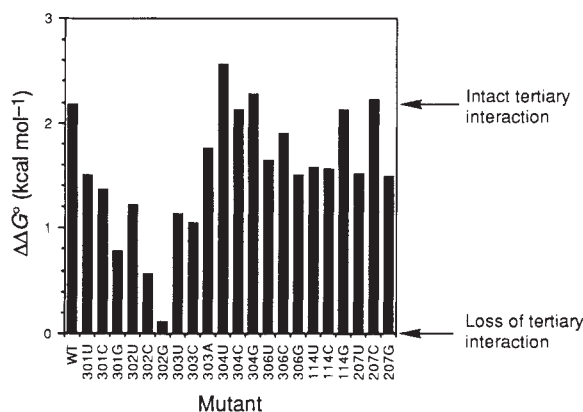


FIG. 2 Ability of WT and mutant ribozymes to distinguish between binding of GGCCCUUCU and GGCCCUd(U)U. $\Delta\Delta G^\circ$ is the difference in ΔG° for binding these two oligonucleotides to a given ribozyme. The majority of mutants, like WT, have a $\Delta\Delta G^\circ$ of ~ 2 kcal mol⁻¹ indicating tighter binding of the oligonucleotide containing the 2'-OH group at -3u. Having a $\Delta\Delta G^\circ$ of nearly zero, A302C and A302G mutants are virtually unable to differentiate, implicating A302 as the likely position of tertiary contact. $\Delta\Delta G^\circ = \ln(K_d(MPd3)/K_d(MP)) \times (0.00198 \text{ kcal mol}^{-1} \text{ K}^{-1} \times 315 \text{ K})$, where $K_d(MPd3)/K_d(MP)$ is taken from column 3 of Table 1. Nucleotides A301, A302, G303, A304, A306, A114 and A207 on the L-21 *Scal* form of the *Tetrahymena* ribozyme² were individually mutagenized to give all three possible single-base mutations³⁸. Binding was measured using a gel-shift assay^{12,15} at constant 42 °C and 10 mM Mg^{2+} . The gel contained 29:1 acrylamide:bisacrylamide and 1 $\times$ electrophoresis buffer (66 mM HEPES, 34 mM Tris base, 0.1 mM EDTA, 10 mM Mg^{2+} , pH 7.5). Samples were equilibrated and loaded in a buffer of 50 mM Tris, 0.1 mM EDTA, 10 mM NaCl, 10 mM $MgCl_2$, 3% glycerol and 0.05% xylene cyanol at pH 7.5. Gels were loaded while running at 10 W in the electrophoresis buffer described above.

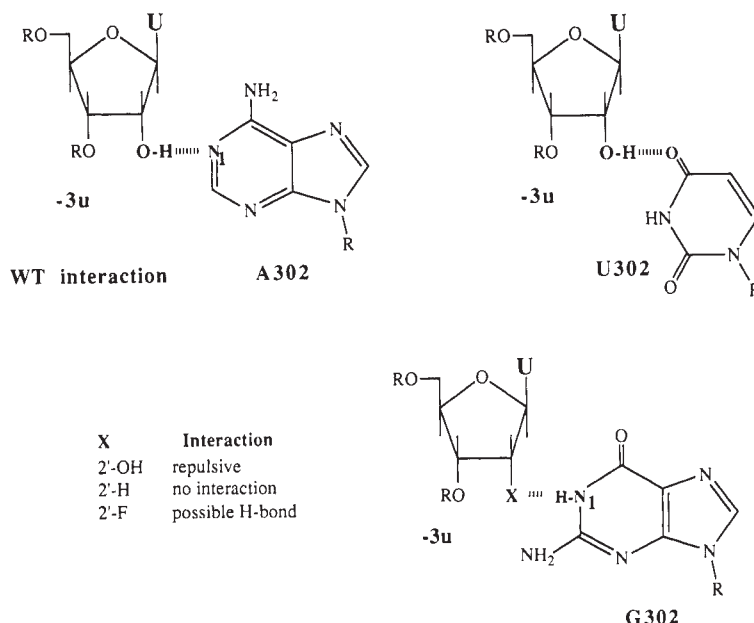


FIG. 3 Model for the WT ribozyme-substrate tertiary interaction (top left) and the A302U-substrate tertiary interaction (top right). In the predictions for interactions with the A302G mutant (bottom), the negative interaction for X=2'-OH is taken as an indication that the 2'-OH is not free to rotate and accept a hydrogen-bond from the N1 functionality of G302 (see text). The 2'-F substituent, which has been used previously in studies of the hammerhead ribozyme^{39,40}, is a potential hydrogen-bond acceptor⁴¹⁻⁴⁵, although the energetic contribution of such a hydrogen-bond has not been determined.

the contact is removed by alteration of the oligonucleotide, where K_d increases by a roughly similar amount, from 0.3 nM (for MP) to 10 nM (for MPd3). Note that these binding energies are still tighter than base-pairing alone (calculated from nearest-neighbour energies to be 600 nM at 42 °C)²¹, so equivalence in binding of MP and MPd3 by the mutant ribozymes cannot be explained by disruption of all sources of tertiary stabilization.

For each of the mutants at the 302 position, binding of MPd2 is at least threefold worse than binding of MP. Furthermore, these are the only mutants that show uniformly poorer binding of MPd2 than of MPd3 (Table 1, columns 4 and 5). These results indicate that the contact to the -2c 2'-OH is still in place, even when the contact to the -3u 2'-OH has been eliminated. Thus, low MP/MPd3 differentiation by the A302N mutants is unlikely to result from indirect effects on ribozyme conformation. Furthermore, this result shows that removal of one interaction does not prevent the other from forming, providing additional evidence for independence of the two tertiary contacts¹⁵.

A contact to the -2c 2'-OH was not identified. Given a -3u contact to A302 and the proposed orientation of P1 relative to the core^{17,20}, the -2c 2'-OH might be in a position to contact

G303. WT ribozyme differentiates MP from MPd2 by a factor of only 6, so although G303C and G303U showed less ability to differentiate MP from MPd2 than most of the mutants, the significance of these effects is difficult to assess. It is possible that the -2c 2'-OH contact might involve the backbone rather than a base of the ribozyme.

To model the interaction between -3u and A302, it was necessary to assign the 2'-OH group as a hydrogen-bond donor or an acceptor. This assignment was based on two lines of reasoning. First of all, direct contact from a substrate 2'-OH group to A302 or U302 can be rationalized if the contact involves N1 of adenine and O4 of uracil acting as hydrogen-bond acceptors. Although adenine and uracil do not share a common pattern of hydrogen bonding, there is exactly 4.0 Å from adenine N1 and uracil O4 to their glycosidic nitrogen atoms. Other pairs of functional groups were disqualified because they were common to G or C, or they did not meet the matching distance requirement. Owing to differences in orientation of the two bases, some change in the backbone conformation (or an intervening water molecule) would be required for the O4 of U302 to reach exactly as far as the N1 of A302, but this may explain the fact that U

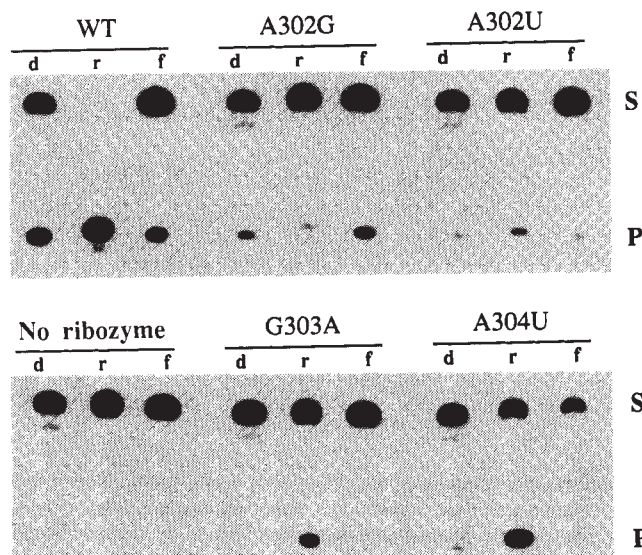


FIG. 4 Specificity of WT and mutant ribozymes for cleavage of 2'-functionalized substrates. A denaturing polyacrylamide gel separates the 5' ³²P-labelled product (P) from substrate (S). Time-points were taken after 6 h of reaction, under $(k_{cat}/K_m)^S$ conditions, as described in legend to Table 1. $(k_{cat}/K_m)^S$ represents the free energy difference between free S and ribozyme, bound with G, and the transition state for the chemical cleavage step. Subheadings d, r and f represent the oligonucleotides dMS, dMSr3 and dMSf3, respectively, as defined in the text. WT and all of the mutants except A302G preferentially cleave dMSr3. A302G prefers dMSf3 instead. A302C (not shown) is relatively unreactive, making it difficult to evaluate its specificity.

does not fully substitute for A at position 302. The second argument for the 2'-OH serving as a proton donor comes from WT ribozyme cleavage experiments using a single 2'-F substituent at the -3 position on a DNA substrate. The fluorinated substrate reacts slower than an all-DNA substrate. This suggests that the electronegative 2'-F, a possible hydrogen-bond acceptor, is clashing with another electronegative hydrogen-bond acceptor on the WT ribozyme (D. Herschlag, F. Eckstein and T.R.C., unpublished results). A model based on these lines of reasoning involves the 2'-OH of -3u donating a hydrogen-bond to the N1 of adenine 302 in the ribozyme core (Fig. 3).

In the A302G mutant, one might expect that the protonated N1 group of G302 would be able to donate a hydrogen-bond to the 2'-oxygen atom at -3u. The failure of A302G to recognize this 2'-OH group is consistent with the idea that the -3u 2'-OH only donates a hydrogen bond to position 302 in the context of this ribozyme. For example, the 2'-oxygen might not accept a hydrogen-bond from G302 because it is already involved as an acceptor in another tertiary interaction. Although it is always possible that the exocyclic amino group of G302 might sterically hinder RNA substrate binding, the A302C ribozyme, which has a smaller base at position 302, has similar binding properties (Table 1).

Tertiary interaction specificity in catalysis

The 2'-OH at the -3u position of substrate contributes to transition state stabilization (D. Herschlag and T.R.C., unpublished results). In the cleavage of DNA substrates by the WT ribozyme, the kinetic parameter $(k_{\text{cat}}/K_m)^S$ incorporates both the binding step and the chemical step³. When a 2'-OH was substituted at the -3 position of a DNA substrate, $(k_{\text{cat}}/K_m)^S$ increased by 13-fold whereas substitution of a 2'-F at this position decreased $(k_{\text{cat}}/K_m)^S$ by 2.5-fold relative to the all-deoxy substrate (our 42 °C data are consistent with earlier data at 50 °C and 30 °C by D. Herschlag, F. Eckstein and T.R.C., unpublished results).

To test the model (Fig. 3) for tertiary interaction in the transition state, the mutant ribozymes were evaluated for their ability to cleave the dMSr3, dMSf3, and dMS oligonucleotide substrates (d(CCCr(U)CUA₅), d(CCC2'F(U)CUA₅) and d(CCCUCUA₅), respectively; Table 1, columns 6 and 7; Fig. 4). The mutants at positions other than 302, like the wild-type ribozyme, showed strong preference for cleavage of dMSr3 and, for the most part, a twofold preference for cleavage of dMS over dMSf3 (data not shown). A302G was unique among mutants studied in that it cleaved dMS preferentially to dMSr3. Because the k_{cat}/K_m preference is opposite to that of the WT ribozyme, it constitutes by definition a reversal of specificity.

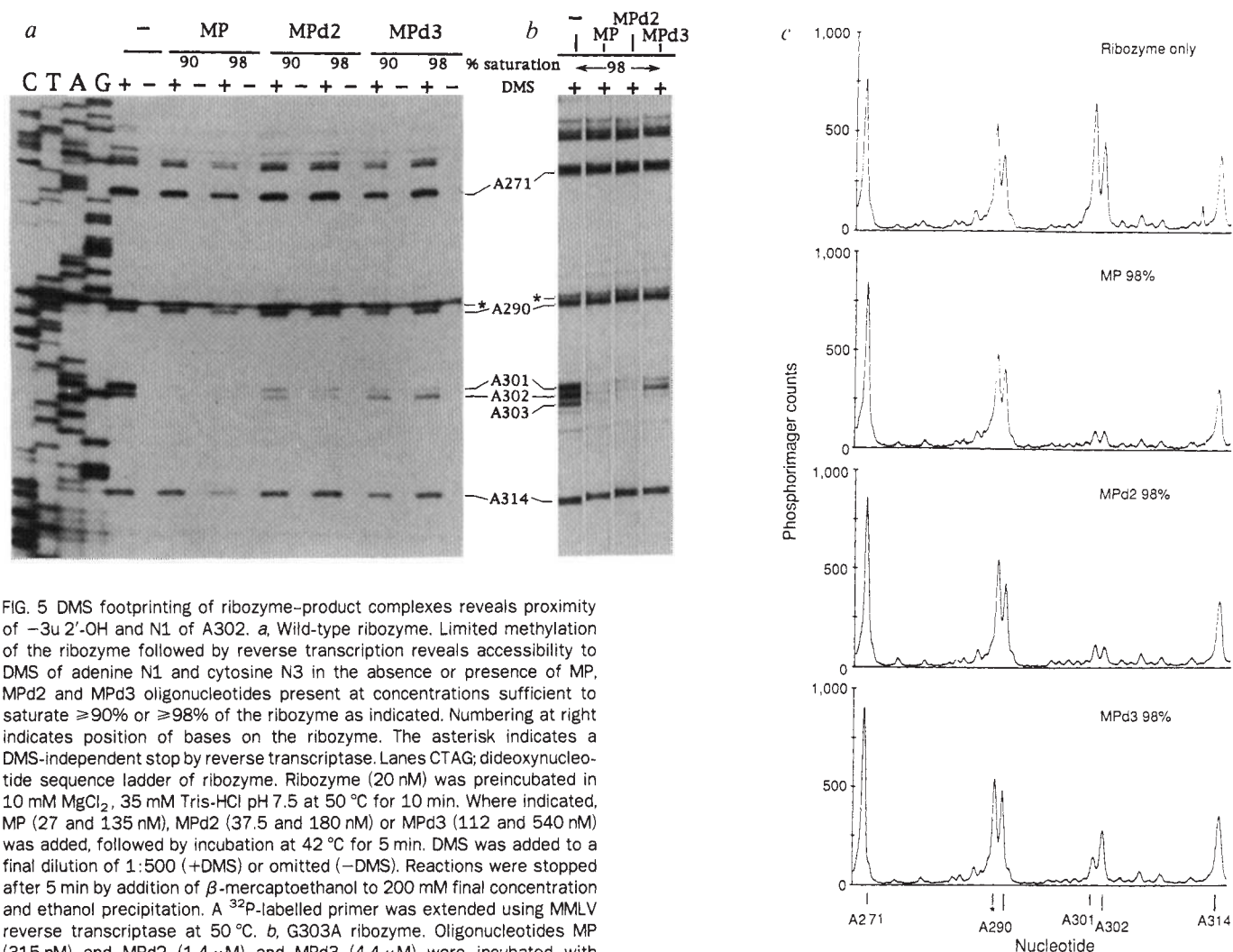


FIG. 5 DMS footprinting of ribozyme-product complexes reveals proximity of -3u 2'-OH and N1 of A302. *a*, Wild-type ribozyme. Limited methylation of the ribozyme followed by reverse transcription reveals accessibility to DMS of adenine N1 and cytosine N3 in the absence or presence of MP, MPd2 and MPd3 oligonucleotides present at concentrations sufficient to saturate $\geq 90\%$ or $\geq 98\%$ of the ribozyme as indicated. Numbering at right indicates position of bases on the ribozyme. The asterisk indicates a DMS-independent stop by reverse transcriptase. Lanes CTAG; dideoxynucleotide sequence ladder of ribozyme. Ribozyme (20 nM) was preincubated in 10 mM MgCl₂, 35 mM Tris-HCl pH 7.5 at 50 °C for 10 min. Where indicated, MP (27 and 135 nM), MPd2 (37.5 and 180 nM) or MPd3 (112 and 540 nM) was added, followed by incubation at 42 °C for 5 min. DMS was added to a final dilution of 1:500 (+DMS) or omitted (-DMS). Reactions were stopped after 5 min by addition of β -mercaptoethanol to 200 mM final concentration and ethanol precipitation. A ³²P-labelled primer was extended using MMLV reverse transcriptase at 50 °C. *b*, G303A ribozyme. Oligonucleotides MP (315 nM) and MPd2 (1.4 μ M) and MPd3 (4.4 μ M) were incubated with ribozyme and treated as described above. *c*, Quantitation of select lanes from *a*, using a Phosphorimager (Molecular Dynamics). Relative to the methylation of free ribozyme, the fractions of methylation at position 301 in the presence of 98% saturating MP, MPd2 and MPd3 are 0.11, 0.14 and 0.19, respectively; at position 302, the fractions are 0.18, 0.21 and 0.64,

respectively. Quantification of this experiment had a precision of $\pm 10\%$. The qualitative trend for the wild-type ribozyme was confirmed in three independent sample preparations, each represented by two methylation timepoints.

A302G also showed a 50% preference for cleavage of dMSf3 over dMS, again opposite to the WT ribozyme. These kinetic results show that A302 contributes to the tertiary interaction in the transition state as well as the ground state. Because the 2'-OH of -3u participates in the interaction in both ground state and transition state, it is necessary that its pairing partner, A302, do so as well.

These results are in accord with our model of -3u 2'-OH-N1 A302 contact as follows. WT and the majority of mutant ribozymes should be able to make the tertiary contact and preferentially cleave dMSr3, but have a deleterious interaction with the electronegative 2'-F of dMSf3. In the A302G mutant, N1 of G at position 302 is protonated, giving a bad interaction with dMSr3. When the 2'-H or 2'-F is substituted, this negative interaction is relieved. The preference for 2'-F relative to 2'-H might indicate that the fluorine accepts a hydrogen-bond from N1 of G, or it could result from more stable base-pairing because of the 2'-fluoro substitution^{22,23}. In fact, cleavage of dMSf3 has a decreased K_m relative to that of the other oligonucleotide substrates for both WT and A302G ribozymes (data not shown). Because deoxynucleotide substitution has no such stabilizing effect, in some cases even destabilizing base-pairing^{15,16}, the preference for dMS over dMSr3 with the A302G ribozyme cannot be attributed to more favourable base-pairing interactions. In summary, the cleavage data support the model shown in Fig. 3.

Footprint of the tertiary contact

A direct probe for the proposed interaction between N1 of A302 and -3u 2'-OH in the wild-type ribozyme was provided by dimethyl sulphate (DMS) modification^{24,25}. Adenine N1 functionalities in a folded RNA molecule that are not protected by base-pairing, tertiary interactions, or steric effects are methylated by DMS. In the absence of RNA substrate, A301 and A302 were readily methylated by DMS (Fig. 5a). In the presence of saturating MP or MPd2, these sites became protected from methylation, and no other sites of protection were observed. Most striking is that, in the presence of saturating MPd3, specific methylation

of A302 was restored to 60% of that seen with free ribozyme, whereas A301 remained protected (Fig. 5a, c). This specific deprotection of A302 N1 on a modest functional group exchange (a 2'-H replacing a 2'-OH at -3u) indicates that A302 N1 and the -3u 2'-OH are in communication and is consistent with a direct interaction between these two functional groups. Loss of A302 protection might occur through increased flexibility on loss of the stabilizing bond. Although this type of result is often taken as proof of a direct interaction, it is possible that N1 exposure is the indirect consequence of loss of a stabilizing interaction elsewhere.

To insure further that interaction with N1 is specific to the 302 position, methylation footprinting was done on the G303A mutant ribozyme. This ribozyme allows N1 functionalities on both sides of A302 to be visualized. Adenines at positions 301, 302 and 303 were protected from DMS in the presence of MP and MPd2, whereas protection at position 302 was specifically reduced in the presence of MPd3 (Fig. 5b). Thus, N1 of A302 seems to be the site of specific interaction with bound oligonucleotide, whereas A301 and (in the mutant) A303 are either sterically protected or become involved in tertiary interactions when P1 binds to the core.

Discussion

The question of how the 5'-splice site is held into place for cleavage by GTP has been a long-standing one. The demonstration that the nucleotides preceding the 5'-splice site are base-paired to a portion of the internal guide sequence (IGS) to form the P1 helix gave a partial answer^{1,6,7}, but led to the question of how the P1 helix was held into place. We have now presented evidence that a tertiary interaction between a specific base in the ribozyme core and a 2'-OH group on the substrate-IGS helix contributes to P1 binding. Thermodynamic measurements, reaction kinetics data and chemical probing experiments are all consistent with a hydrogen-bond involving donation of a proton from the 2'-OH group three nucleotides from the cleavage site on the substrate to the N1 functional group on A302 in the phylogenetically conserved ribozyme core (Fig. 6a).

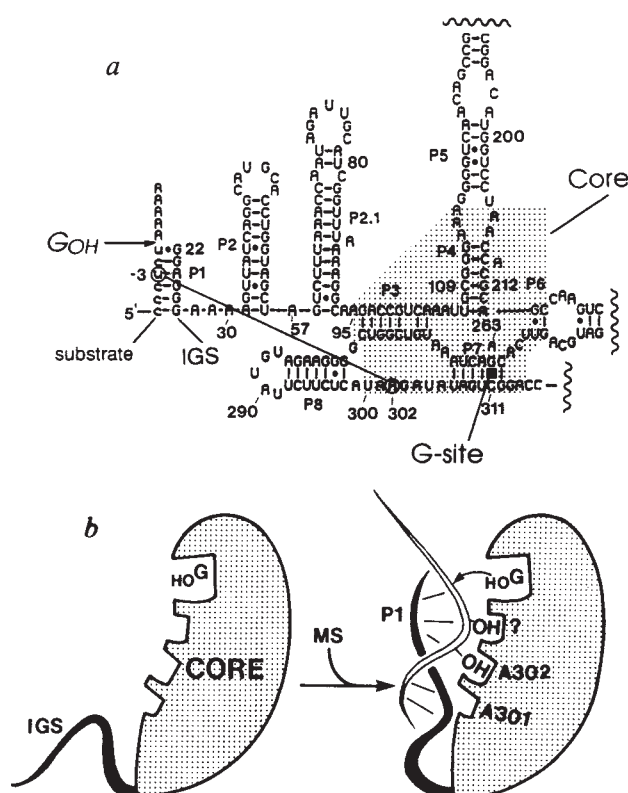


FIG. 6 a, Secondary structure of the *Tetrahymena* ribozyme. Dark line indicates the contact between the substrate 2'-OH located three nucleotides from the cleavage site and the ribozyme residue A302. Such a contact would position the P1 helix near the G-binding site indicated at the 264-311 base-pair. The catalytic core of the ribozyme is shaded and regions corresponding to continuations of the ribozyme outside the core have been abbreviated with wavy lines. b, Representation of the *Tetrahymena* ribozyme in its free and substrate-bound forms. In the free form, the internal guide sequence (IGS) may be conformationally unrestricted or bound in an alternative manner that leaves A301 and A302 accessible. On binding of substrate (white strand) and formation of P1, the IGS becomes part of a docked helix, leading to protection of A301 and, if the substrate has a -3u 2'-OH, protection of A302. As described in the text, there is evidence for multiple interactions between P1 and the core in addition to the -3u 2'-OH to N1(A302) contact.

Hydrogen-bonding between non-helical 2'-OH groups and nucleic acid bases has been deduced from crystallographic and nuclear magnetic resonance (NMR) data of RNA molecules²⁶⁻²⁹. Investigations with the *Tetrahymena* ribozyme show that contacts from 2'-OH groups within a helix can make an energetic contribution to RNA structure, recognition and reactivity. The interaction identified in this study strongly stabilizes binding between a ribozyme and its oligonucleotide substrate and contributes to transition state stabilization energy as well. Therefore, 2'-OH groups are not restricted to influencing RNA structure through indirect effects on sugar pucker and helical conformational rigidity, but can also contribute directly to RNA structural complexity by providing additional points of hydrogen-bonding. We expect that this type of interaction is common in structured RNA molecules, because once the bases of an RNA molecule have packed into secondary structure, its backbone provides the major surface available for tertiary structure formation. Considering that the RNA subunit of RNase P recognizes a helical region within many different transfer RNA precursors, it is likely that 2'-OH mediated tertiary interactions that are not base-specific contribute to the stability of RNase P-tRNA interactions^{30,31}.

Identification of the tertiary contact provides spatial information necessary to elucidate the three-dimensional architecture of group I introns. The contact we have identified helps to stabilize and orient P1 in an appropriate position for interaction with the guanosine nucleophile⁸ and the catalytic apparatus (Fig. 6a). A three-dimensional model for the structure involved in the second step of group I RNA self-splicing was proposed by Michel and Westhof²⁰. The overall architecture of their model agrees well with our data. On the basis of their model they suggested that G303 could contact the ribose sugar at -3u, which does not agree with our analysis. But a local revision of their model can incorporate the interaction identified here (F. Michel and E. Westhof, personal communication). The use of phylogenetic/molecular modelling, together with experimental identification of tertiary interactions, provides a particularly powerful approach to the understanding of RNA structure.

The modification footprinting suggests that a substrate-induced conformational change may occur in the ribozyme structure. Residue 301 in the core was vulnerable to DMS attack in the absence of RNA substrate and protected in the presence of RNA substrate, but unlike A302, its protection was not dependent on the presence of a 2'-OH at the -3 position of substrate. Our interpretation is that protection of A301 occurs when P1 docks into the core (Fig. 6b). As the P1 helix turns,

substrate 2'-OH groups become inaccessible to the face of the ribozyme core, but 2'-OH groups on the IGS are poised for interaction with core base residues below A302, such as A301. In their revised model of the group I intron core mentioned above, Michel and Westhof (personal communication) postulate a contact between A301 and the 2'-OH of G25 on the IGS; such a contact is fully consistent with our results. Independent of the molecular identity of this contact, our data indicate that it is not present in the free ribozyme, but occurs only when P1 is formed. A kinetic description of a pathway for P1 docking, including an induced-fit conformational change, has recently been proposed³². Additional evidence for such a conformational change has been obtained by derivatization of G22 in the IGS with a cross-linking reagent. These data indicate that G22, which pairs with -1u at the cleavage site in P1, is near A114 in the core only when P1 forms (J.-F. Wang, W. Downs and T.R.C., unpublished results).

The all-deoxy MP, d(CCCUCU), also confers partial protection of A301 and A302 from DMS (F.L.M. and T.R.C., unpublished results). This suggests that there are additional tertiary interactions between P1 and the core that are not mediated by 2'-OH groups on the product. The G-U pair at the cleavage site in P1 is phylogenetically conserved and is a candidate for tertiary interaction either by recognition of its geometry³³ or bonding with specific functional groups³⁴. It is therefore likely that multiple tertiary interactions stabilize the docking of P1 into the core of the ribozyme. This may at first seem inconsistent with our finding that tertiary interactions to oligoribonucleotide 2'-OH groups plus base-pairing energy account for most or all of the ribozyme-oligoribonucleotide association energy¹⁵. But the structural reorganization required for P1 docking may be energetically unfavourable, and additional tertiary interactions might provide the energetic compensation necessary to promote binding. In addition, some tertiary interactions might occur only in the transition state and therefore be invisible in the ground state.

We know from previous work that the sugar-phosphate backbone of the core is protected from attack by hydroxyl radicals^{35,36}. The core of the ribozyme may therefore be a tightly packed environment, containing highly ordered water molecules, and there might be substantial energetic benefit to hydrogen-bonding from 2'-OH groups. Like base-pairing and base-triple interactions, 2'-OH groups can bring one strand of nucleic acid close to another. In this respect, the sugar-base interactions elucidated in this work are representative of a third form of RNA-RNA recognition. □

Received 7 May; accepted 9 June 1992.

1. Zaug, A. J., Been, M. D. & Cech, T. R. *Nature* **324**, 429-433 (1986).
2. Zaug, A. J., Grosshans, C. A. & Cech, T. R. *Biochemistry* **27**, 8924-8930 (1988).
3. Herschlag, D. & Cech, T. R. *Nature* **344**, 405-409 (1990).
4. Davies, R. W., Waring, R. B., Ray, J. A., Brown, T. A. & Sczozocchio, C. *Nature* **300**, 719-724 (1982).
5. Michel, F., Jacquier, A. & Dujon, B. *Biochimie* **65**, 867-881 (1982).
6. Been, M. D. & Cech, T. R. *Cell* **47**, 207-216 (1986).
7. Waring, R. B., Townner, P., Minter, S. J. & Davies, R. W. *Nature* **321**, 133-139 (1986).
8. Michel, F., Hanna, M., Green, R., Bartel, D. P. & Szostak, J. W. *Nature* **342**, 391-395 (1989).
9. Sullivan, F. X. & Cech, T. R. *Cell* **42**, 639-648 (1985).
10. Sugimoto, N., Kierzek, R. & Turner, D. H. *Biochemistry* **27**, 6384-6392 (1988).
11. Sugimoto, N., Sasaki, M., Kierzek, R. & Turner, D. H. *Chem. Lett.* 2223-2226 (1989).
12. Pyle, A. M., McSwiggen, J. A. & Cech, T. R. *Proc. natn. Acad. Sci. U.S.A.* **87**, 8187-8191 (1990).
13. Herschlag, D. & Cech, T. R. *Biochemistry* **29**, 10159-10171 (1990).
14. Doudna, J. A. & Szostak, J. W. *Nature* **339**, 519-522 (1989).
15. Pyle, A. M. & Cech, T. R. *Nature* **350**, 628-631 (1991).
16. Bevilacqua, P. C. & Turner, D. H. *Biochemistry* **30**, 10632-10640 (1991).
17. Kim, S.-H. & Cech, T. R. *Proc. natn. Acad. Sci. U.S.A.* **84**, 8788-8792 (1987).
18. Michel, F., Ellington, A. D., Couture, S. & Szostak, J. W. *Nature* **347**, 578-580 (1990).
19. Williamson, C. L., Tierney, W. M., Kerker, B. J. & Burke, J. M. *J. biol. Chem.* **262**, 14672-14682 (1987).
20. Michel, F. & Westhof, E. *J. molec. Biol.* **216**, 585-610 (1990).
21. Freier, S. M. et al. *Proc. natn. Acad. Sci. U.S.A.* **83**, 9373-9377 (1986).
22. Guschlbauer, W. & Jankowski, K. *Nucleic Acids Res.* **8**, 1421-1433 (1980).
23. Janik, B. et al. *Biochem. biophys. Res. Commun.* **46**, 1153-1160 (1972).
24. Inoue, T. & Cech, T. R. *Proc. natn. Acad. Sci. U.S.A.* **82**, 648-652 (1985).
25. Moazed, D., Stern, S. & Noller, H. F. *J. molec. Biol.* **187**, 399-416 (1986).
26. Quigley, G. J. & Rich, A. *Science* **194**, 796-806 (1976).

27. Dock-Bregeon, A. C. et al. *J. molec. Biol.* **209**, 459-474 (1989).
28. Holbrook, S. R., Cheong, C., Tinoco, I. & Kim, S.-H. *Nature* **353**, 579-581 (1991).
29. Heus, H. A. & Pardi, A. *Science* **253**, 191-194 (1991).
30. Pace, N. R. & Smith, D. *J. biol. Chem.* **265**, 3587-3590 (1990).
31. Forster, A. C. & Altman, S. *Science* **249**, 783-786 (1990).
32. Herschlag, D. *Biochemistry* **31**, 1386-1399 (1992).
33. Doudna, J. A., Cormack, B. P. & Szostak, J. W. *Proc. natn. Acad. Sci. U.S.A.* **86**, 7402-7406 (1989).
34. Barford, E. T. & Cech, T. R. *Molec. cell. Biol.* **9**, 3657-3666 (1989).
35. Latham, J. A. & Cech, T. R. *Science* **245**, 276-282 (1989).
36. Celander, D. W. & Cech, T. R. *Science* **251**, 401-407 (1991).
37. Weeks, K. M. & Crothers, D. M. *Cell* **66**, 577-588 (1991).
38. Kunkel, T. A., Roberts, J. D. & Zakour, R. A. *Meth. Enzym.* **154**, 367-382 (1987).
39. Pieken, W. A., Olsen, D. B., Benseler, F., Aarup, H. & Eckstein, F. *Science* **253**, 314-317 (1991).
40. Williams, D. M., Pieken, W. A. & Eckstein, F. *Proc. natn. Acad. Sci. U.S.A.* **89**, 918-921 (1992).
41. Curtiss, L. A., Frurip, D. J. & Blander, M. *J. Am. chem. Soc.* **100**, 79-86 (1978).
42. Murray-Rust, P., Stallings, W. C., Monti, C. T., Preston, R. K. & Glusker, J. P. *J. Am. chem. Soc.* **105**, 3206-3214 (1983).
43. Withers, S. G., Street, I. P. & Percival, M. D. *ACS Symp. Series* **374**, 59-78 (1988).
44. Takahashi, L. H., Radhakrishnan, R., Rosenfeld, R. E., Meyer, E. F. & Trainor, D. A. *J. Am. chem. Soc.* **111**, 3368-3374 (1989).
45. Huang, J. & Hedberg, K. *J. Am. chem. Soc.* **111**, 6909-6913 (1989).

ACKNOWLEDGEMENTS. We thank F. Eckstein for 2'-fluorouridine amidite, A. Gooding for synthesis of oligonucleotides containing 2'-fluorouridine, D. Herschlag for discussions and for permission to cite unpublished results, J. Piccirilli for discussions on modelling the tertiary interaction and F. Michel and E. Westhof for information and discussion. This work was supported by the Jane Coffin Childs Fund for Medical Research (A.M.P.), the National Institutes of Health (T.R.C.) and the W. M. Keck Foundation.

High lithium abundance in the secondary of the black-hole binary system V404 Cyg

E. L. Martín, R. Rebolo, J. Casares & P. A. Charles*†

Instituto de Astrofísica de Canarias, 38200 La Laguna, Tenerife, Spain

* Royal Greenwich Observatory, Apartado 321,

38780 Santa Cruz de La Palma, Canary Islands, Spain

† Department of Astrophysics, Nuclear Physics Building, Keble Road, Oxford OX1 3RH, UK

THE maximum abundance of lithium in the oldest stars of the Galaxy is a factor of ten less than that of young stars and the local interstellar medium¹, prompting an active search for sources of lithium. Type II supernovae, novae and accretion disks around black holes may be important sites of lithium production^{2,3}. Stars in general are unlikely to produce lithium, as it is destroyed in their interiors through (p, α) reactions. Recently we showed that the transient X-ray source V404 Cyg is a low-mass X-ray binary with a black-hole primary⁴. Here we describe the discovery of lithium in the spectrum of V404 Cyg, as has also been noted independently by Wallerstein⁵ in our previously published spectra⁴. We derive a high lithium abundance in the secondary star of V404 Cyg, close to that of very young stars. Without lithium production, the secondary would have to be comparable in age to the Pleiades cluster (~ 100 Myr). If we are not seeing the system at an early moment in its lifetime, there must have been lithium production, which could be associated either with the supernova explosion that created it, or with the accretion disk now there. Further observations of this and similar systems may reveal that they are important sources of galactic lithium enrichment.

X-ray transient sources such as V404 Cyg are thought to be close binary systems composed of a neutron star or black hole primary accreting mass from a low-mass secondary. X-ray outbursts come from sudden instabilities in either the accretion disk or the donor star. After its 1989 outburst, V404 Cyg has declined to its quiescent level, and the secondary ($V \approx 18.5$), is now optically detectable. Its radial velocity curve⁴ provides compelling evidence for a compact primary of mass ≥ 6 solar masses ($M_{\odot}$), exceeding the $3M_{\odot}$ maximum permitted mass for a neutron star⁶ and making V404 Cyg one of the most secure

of black-hole candidates. During the analysis of the 1991 database subset used for the radial velocity study⁴, consisting of 55 spectra taken at the 4.2-m William Herschel Telescope, we discovered in the spectrum of V404 Cyg the presence of the Li I resonance line at wavelength 6,707.8 Å. In Fig. 1, the sum of all the spectra, the Li I line is clearly seen. We also plot the result of subtracting the spectrum of the G9V standard HR1232, illustrating that the absorption line strengths of the standard and of the secondary of V404 Cyg⁴ are very similar. The Li I absorption line stands out because of its unusual strength in V404 Cyg. We confirmed the detection of the lithium line in various independent subsets of the data sample. In Fig. 2 we focus on the Li region of V404 Cyg for three consecutive nights. Because the orbital period is 6.47 days, each observing night (5–6 hours) corresponds to a different orbital phase. The Li I line is present irrespective of the phase and is modulated in wavelength with the same radial velocity curve as the other secondary star features. Finally, we plot the spectrum of the post-T Tauri star HD33802B, which has a high lithium abundance⁷. The comparison of the spectra of V404 Cyg with that of HD33802B shows that the Li I line-strength in the secondary of the black hole is similar to that of an extremely young star (~ 30 Myr) of similar spectral type.

To calculate chemical abundances for the secondary of V404 Cyg, we estimate the basic parameters of its atmosphere. Using a conversion⁸ of spectral type to effective temperature T_{eff} we find $5,250 \pm 100$ K for a spectral type of G9(± 1)V, which is adequate for a dwarf star of $\log g \approx 4.5$. The secondary of V404 Cyg is more likely to be a dwarf than a giant: the higher luminosity of a giant would place the system at ~ 11 kpc, implying a spectacular X-ray luminosity at outburst^{4,5}, and the corotational velocity expected for a giant is 85 km s^{-1} , much higher than allowed by the rotational broadening of the secondary (J.C. *et al.*, manuscript in preparation). Nevertheless, we also calculated the abundances for the case where the secondary is a KOIII star, with $T_{\text{eff}} 4,655$ K (ref. 8) and $\log g = 3.0$. In all the computations we assumed a microturbulent velocity of 2 km s^{-1} . The metallic lines used were selected because they are relatively strong, unblended and not too far from the Li I line. We estimated that, in the spectral range of interest (6,400–6,800 Å), the veiling effect (J.C. *et al.*, manuscript in preparation) due to the accretion disk was lower than the uncertainty in the measurements and we therefore applied no correction. Table 1 shows the resulting abundances for a dwarf secondary. The calculations

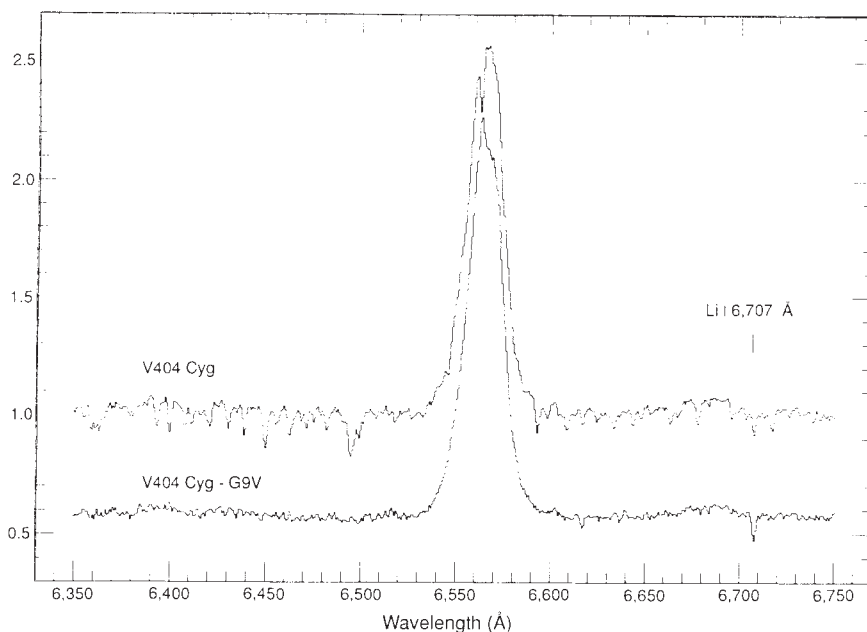


FIG. 1 The sum of all 55 intermediate resolution spectrograms (0.74 Å per pixel) of V404 Cyg taken at the 4.2-m William Herschel Telescope in July and August 1991. Shifted downwards is a plot of the result of subtracting the spectrum of the G9V standard HR1232 from that of V404 Cyg. The only spectral features that remain after subtraction are H α in emission and Li I in absorption.

TABLE 1 Chemical abundances for V404 Cyg

Line wavelength (Å)	log <i>gf</i>	Ref.	<i>W</i> _{obs} (m Å)	Abundance
Fe I λ 6,430.9	-1.99	22	260 $\pm$ 50	0.5 $\pm$ 0.3
Ca I λ 6,439.1	0.47	23	280 $\pm$ 40	0.1 $\pm$ 0.3
Ni I λ 6,643.6	-2.79	24	190 $\pm$ 50	0.5 $\pm$ 0.4
Li I λ 6,707.8	-1.48	23	290 $\pm$ 50	3.2 $\pm$ 0.3
Ca I λ 6,717.7	0.47	23	250 $\pm$ 50	0.4 $\pm$ 0.3

Note: abundance for lithium is $12 + \log (N_{\text{Li}}/N_{\text{H}})$; abundances for metals are $\log (X/H)_{\text{star}} - \log (X/H)_{\text{Sun}}$. *gf* is oscillator strength; *W* is equivalent width.

were made with our local thermodynamic equilibrium (LTE) spectral synthesis code⁷, except for the lithium abundance, which was calculated in non-LTE from curves in the literature⁹. The lithium abundance obtained with our LTE code is a factor of 2 higher than the non-LTE value adopted here, because the Li I line saturates. The error bars in the abundances come from the combined effect of uncertainties in the measured equivalent widths and temperature. The metal abundances seem to be higher than solar values by a factor of 2.5, and the lithium abundance is close to the cosmic interstellar value ($\log N(\text{Li}) \approx 3.1$, where *N* is number density, in the standard scale of $\log N(\text{H}) = 12$). Such a high Li abundance is unusual for a main sequence G/K star as Li is very efficiently depleted in this spectral type range¹⁰. If the secondary were a giant star the metal abundances would still be higher than solar by about a

factor of 2 and the non-LTE lithium abundance would be $\log N(\text{Li}) = 2.7$. This abundance is much higher than generally observed among G/K giant stars¹².

The high Li abundance in the secondary of V404 Cyg must mean either that the star has preserved its initial cosmic Li abundance, or that additional Li has been created in the environment of the binary system.

In post-T Tauri stars⁷ of age ~ 30 Myr, Li is cosmic in the spectral type range G2–G8, and underabundant by a factor ~ 10 for spectral types K1–K2. At the age of the α Per open cluster (~ 50 Myr) some G9-type members are found with Li close to cosmic but the spread of abundances among G/K stars is about an order of magnitude¹³; the same is true for the slightly older (~ 100 Myr) Pleiades cluster¹⁴. The abundance of Li in G/K main-sequence stars gradually declines from the age of the Pleiades, through that of Ursa Major¹⁵ (~ 300 Myr), to the age of the Hyades (~ 800 Myr), where the depletion is 2–3 orders of magnitude¹⁶. If the Li abundance of the secondary of V404 Cyg reflects its initial endowment, the star is much younger than the Hyades, perhaps even younger than the Pleiades. It may be argued that for a late-type star in a close binary system, tidal effects¹⁷ could have affected any mixing mechanism which might otherwise cause the depletion of Li from the atmosphere, making it a poor estimator of the age of the system, but there is no observational evidence that tidal interactions might be efficient at inhibiting Li depletion. The close binary system VB22 in the Hyades, which has a similar spectral type and orbital period (G8V, 5.6 days) to the secondary of V404 Cyg, shows Li depletion of half the mean depletion observed in Hyades stars with the same spectral type¹⁸, although Li is still depleted by about a factor of 10. Further work is needed to estimate whether tidal interactions can inhibit Li depletion in late-type stars and particularly in the secondary of V404 Cyg. If all the Li present in this object comes from preservation of the inborn abundance, the age of the Pleiades would be an upper limit to the age of the system.

Alternatively, the high abundance in the secondary of V404 Cyg be due to Li enrichment, which would have occurred early in the life of the secondary, when the precursor of the black hole was a very high mass star ($M/M_{\odot} \approx 40$ –80). This massive star probably ended as a type II supernova or underwent pulsating explosions, both cases representing important sites of lithium generation². The secondary may then have been chemically contaminated. The production of Li in type II supernovae is expected to track the production of heavier elements such as oxygen and iron. The secondary of V404 Cyg seems to have metal abundances higher than solar by a factor of 2–3 suggesting that its Li could have been similarly enhanced. In this case the system must have an age lower than the time required for Li to be depleted by a factor of 10 in a G9V star, that is, a few hundred million years.

Lithium may also be produced by spallation reactions in an accretion torus around the black hole³. The abundance of Li would then depend on the accretion rate, disk viscosity, fraction of mass ejected and mass of the central object. Some of these parameters are very uncertain, making it possible that the Li depleted in the star could be continuously renewed throughout the stellar lifetime, and removing any restriction on the age of the system. A definitive test of this hypothesis may come from γ -ray observations: if most Li comes from spallation reactions there should be strong γ -lines¹⁹ at 478 and 431 keV. Their recent detection in a solar limb flare provides evidence for Li production by interacting α particles in the solar atmosphere²⁰. The best time to look for these lines in V404 Cyg is during the intense X-ray transient events (they seem to occur on timescales of ~ 20 yr). It would be interesting to search for these lines in other X-ray transients to assess how much Li they may produce.

Observations of the Li I resonance line in the secondaries of other low-mass X-ray binaries will enable us to assess the role

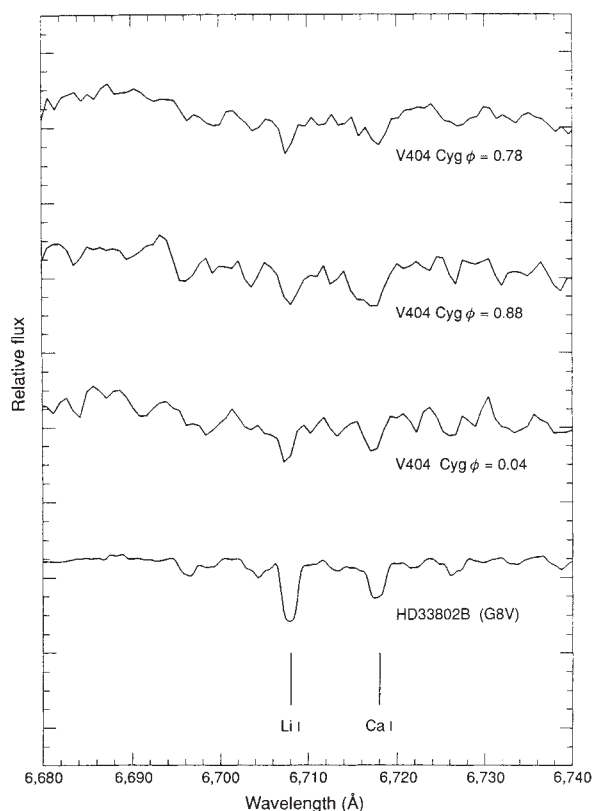


FIG. 2 The expanded Li I spectral region. The three spectrograms of V404 Cyg were taken with the same instrumental configuration (William Herschel Telescope, with ISIS red arm and grating R600) on three different nights (from top downwards, 6, 7 and 8 August 1991) corresponding to three different orbital phases (ϕ). The bottom spectrum of the post-T Tauri star HD33802B was taken with different instrumentation (Isaac Newton Telescope, Intermediate Dispersion Spectrograph, plus grating H1800V) and has been smoothed by an 8-point boxcar to degrade the spectral resolution for visual comparison with V404 Cyg.

of massive compact objects as galactic sources of Li. Conversely the Li abundance could be a powerful probe of the evolution of these systems. The evolutionary timescale of X-ray binaries with a black-hole primary is expected to be much longer ($\sim 1,000$ Myr) than for those with a neutron star primary

(~ 30 Myr)²¹, and therefore comparable to the Li depletion timescale in low-mass stars. If the creation of Li by spallation is comparable to its destruction in the interior of the secondary, detailed modelling will be needed to predict the result of these opposing effects. $\square$

Received 2 March; accepted 18 May 1992.

1. Rebolo, R. in *Evolution of Stars: The Photospheric Abundance Connection*, IAU Symp. 145 (ed. Michaud, G. & Tutukov, A.) 85–97 (1991).
2. Dearborn, D. S. P., Schramm, D. N., Steigman, G. & Truran, J. *Astrophys. J.* **347**, 455–460 (1989).
3. Jin, L. *Astrophys. J.* **356**, 501–506 (1990).
4. Casares, J., Charles, P. A. & Naylor, T. *Nature* **355**, 614–617 (1992).
5. Wallerstein, G. *Nature*, **356**, 569 (1992).
6. McClintock, J. E. & Remillard, R. A. *Astrophys. J.* **308**, 110–122 (1986).
7. Martin, E. L., Magazzù, A. & Rebolo, R. *Astr. Astrophys.* **257**, 186 (1992).
8. de Jager, C. & Nieuwenhuijzen, H. *Astr. Astrophys.* **177**, 217–227 (1987).
9. Magazzù, A., Rebolo, R. & Pavlenko, Ya. V. *Astrophys. J.* (in the press).
10. Herbig, G. H. *Astrophys. J.* **141**, 588–609 (1965).
11. Proffitt, C. R. & Michaud, G. *Astrophys. J.* **346**, 976–982 (1989).
12. Brown, J. A., Sneden, C., Lambert, D. L. & Dutchover, E. Jr. *Astrophys. J. Suppl.* **71**, 293–322 (1989).
13. Balachandran, S., Lambert, D. L. & Stauffer, J. R. *Astrophys. J.* **333**, 267–276 (1988).
14. Duncan, D. K. & Jones, B. F. *Astrophys. J.* **271**, 663–671 (1983).

15. Duncan, D. K. *Astrophys. J.* **248**, 651–669 (1981).
16. Cayrel, R., Cayrel de Strobel, G., Campbell, B. & Däppen, W. *Astrophys. J.* **283**, 205–208 (1984).
17. Zahn, J. P. & Bouchet, L. *Astr. Astrophys.* **223**, 112–118 (1989).
18. Soderblom, D. R., Oey, M. S., Johnson, D. R. H. & Stone, R. P. S. *Astr. J.* **99**, 595–607 (1990).
19. Kozlovsky, B. & Ramaty, R. *Astrophys. J.* **191**, L43–L44 (1974).
20. Murphy, R. J., Hua, X. M., Kozlovsky, B. & Ramaty, R. *Astrophys. J.* **351**, 299–308 (1990).
21. Romani, R. W. in *Aspen Winter Astrophys. Conf. X-ray and γ -ray signatures of Neutron Stars versus Black Holes*, Stanford Univ. preprint (1992).
22. Wiese, W. L., Smith, M. W. & Glennon, B. M. *Atomic Transition Probabilities* NSRDS-NBS 4, Vol. 1 (1966).
23. Gurtovenko, E. A. & Kostik, R. I. *Astr. Astrophys. Suppl.* **46**, 239–248 (1981).
24. Kurucz, R. L. & Peytremann, E. A. *Smithsonian Astrophysical Observatory Rep No.* 362 (1975).

ACKNOWLEDGEMENTS. The William Herschel Telescope is operated by the Royal Greenwich Observatory in the Spanish Observatorio del Roque de los Muchachos of the Instituto de Astrofísica de Canarias.

Ozone depletion in the upper stratosphere estimated from satellite and Space Shuttle data

Ernest Hilsenrath*, Richard P. Cebula†
& Charles H. Jackman*

* NASA/Goddard Space Flight Center, Greenbelt Maryland, 20771, USA

† Hughes STX Corporation, Lanham, Maryland 20706, USA

WHEN the possibility of anthropogenic ozone depletion was first identified, it was thought that it would occur primarily in the upper stratosphere, at altitudes near 42 km. It is now recognized that ozone losses due to heterogeneous reactions involving chlorine and bromine¹ are greatest in the lower stratosphere, near 20 km (ref. 2), and this is the main cause of ozone depletion over polar latitudes³. Despite satellite observation of the upper stratosphere for nearly a decade⁴, the question of possible ozone depletion in this region has remained unresolved because of instrument degradation and incomplete monitoring⁵. Recent observations of ozone concentrations in the upper stratosphere have been made with the Shuttle Solar Backscatter Ultraviolet (SSBUV) spectrometer carried by the Space Shuttle. Here we combine the SSBUV data for October 1989 with measurements made in October 1980 by the similar SBUV instrument on NASA's Nimbus-7 satellite, to show that the ozone concentration near 45 km has decreased during this period by about $7 \pm 2\%$. The trend is consistent with the predictions of a two-dimensional photochemical model. Although this contribution to total column ozone depletion is small, the changes may have implications for the radiative properties of the upper atmosphere.

The SSBUV experiment is designed to provide calibration checks to the series of SBUV/2 instruments (an improved version of the Nimbus-7 SBUV) flying on the NOAA operational satellites for the purpose of detecting global trends in the vertical distribution and the total column amount of ozone. The checks are done through a series of SSBUV flights, where comparisons are made using coincident observations from the Shuttle and satellite instruments^{6–8}. The SSBUV instrument is nearly identical to ozone instruments flying on the Nimbus and NOAA satellites. Its first flight was in October 1989, with subsequent flights in October 1990, August 1991 and March 1992.

The limited ability of the Nimbus-7 SBUV data to detect trends because of instrument degradation has been described in detail⁵, but the data are readily available. In 1979 and 1980,

when the Nimbus-7 SBUV degradation was well characterized⁵, and 1989, when the SSBUV coincidences were done, the solar cycle was at its maximum. Solar irradiance in the middle ultraviolet, the primary source for ozone production, follows the 11-year solar cycle⁹. An analysis of trends over one solar cycle removes the uncertainty, at least to first order, of the effects of solar ultraviolet changes. This allows us to estimate the effects on decadal ozone trends of chlorine-related catalytic processes resulting from anthropogenic gases reaching the stratosphere.

Figure 1 illustrates the orbital crossings of the Shuttle and the Nimbus-7 satellite that locate the measurement coincidences over the $2\frac{1}{2}$ days of SSBUV operation. A coincidence is the location where the instantaneous fields of view of instruments on the Shuttle and satellite are the same to within 1° latitude and longitude on the Earth within a 1-hour period. The solar zenith angles differed by no more than 12° over this period. There were 38 crossings with 76 observations that met the criteria. The 1-hour window is small enough that atmospheric variability does not introduce a significant error into our analysis^{6,8}.

Data are limited to daylight periods only, because both the SSBUV and SBUV observations use the backscattered solar radiation in the ultraviolet to derive ozone amounts. The two bold horizontal lines in Fig. 1 mark the latitude range of SSBUV data. The inclination of the Shuttle orbit establishes the northern boundary, and the evening and morning terminators limit the southern boundary as the Shuttle travels eastwards.

Figure 2 illustrates the comparison of ozone values from SSBUV and the Nimbus-7 SBUV in 1989 for a pressure range

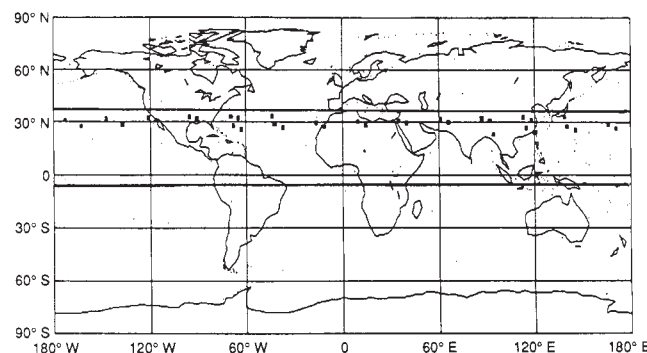


FIG. 1 Location of crossings of SSBUV and Nimbus-7 SBUV. There were 38 crossings and 76 coincidences that met the 1-hour criterion.

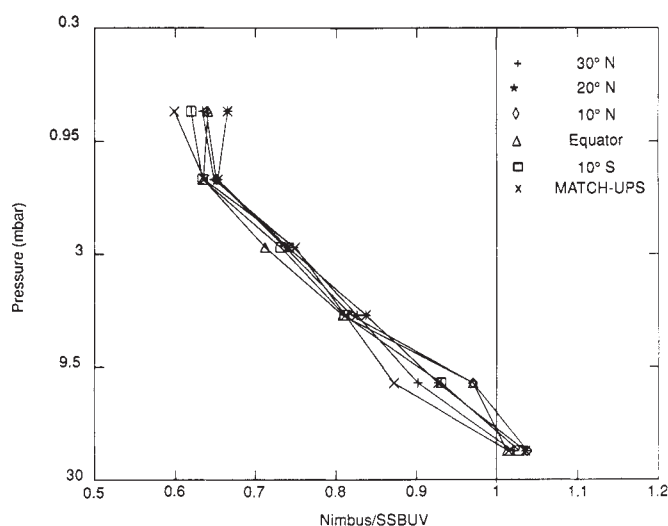


FIG. 2 SSBUV and Nimbus-7 SBUV comparisons are shown as the ratio of the zonal average ozone amounts (10° S to 30° N) measured by Nimbus-7 to those measured by SSBUV over a pressure range corresponding to 25–50 km.

22.4–0.7 mbar (26–51 km altitude). For backscatter ultraviolet measurements, the shorter wavelengths are used to sense higher-altitude ozone. The comparison shows a monotonic decrease of Nimbus-7 values relative to SSBUV values with increasing altitude. Therefore Fig. 2 is consistent with the fact that the Nimbus-7 SBUV degradation over time was more severe at shorter wavelengths⁵. Nimbus ozone values are shown relative to SSBUV values for two cases. For the first case, comparisons consist only of matched data as discussed above. For the second case, comparisons are for zonal averages in 10° latitude bins. Each bin consists of ~ 60 observations from each instrument. For these averages, SSBUV covers a range of larger solar zenith angles (38 – 85°) over the observed latitude range than SBUV (0 – 45°) because the orbital inclination of the Shuttle is lower than that of Nimbus-7. Also, the Nimbus-7 orbit is Sun-synchronous, so the SBUV instrument observes ozone at a near-constant local time, whereas the near-equatorial Shuttle orbit results in SSBUV observing ozone over a wider range of local times. The agreement between the two cases is very good except at 10 and 0.7 mbar, where the difference at these levels is at the limits of the standard errors of the average ($\pm 1\%$). The

differences at these two levels are algorithmic problems related to large solar zenith angles and possible diurnal variations, respectively.

We can now correct for the effect of Nimbus-7 SBUV degradation on the ozone amounts using a simple correction factor derived from Fig. 2. The correction factor, c , shown in equation (1), is a function of pressure height, p , and latitude, l , where O_s and O_n are ozone values from the Shuttle and Nimbus, respectively, as a function of pressure and latitude.

$$c(p, l) = O_s(p, l) / O_n(p, l) \quad (1)$$

The correction factor at each latitude is averaged with the match-up data and then used to correct the zonal average ozone between 30° N and 30° S. The correction was extended to 30° S because of the latitude symmetry of ozone over the tropics. The data are further smoothed by averaging over latitude. An ozone trend is then derived by subtracting 1980 Nimbus-7 data, also averaged over latitude for October, from the corrected Nimbus-7 1989 data. The 1980 data are used because that year matches the phase of the quasi-biennial oscillation. In addition, the Nimbus-7 degradation was properly accounted for in 1980 (ref. 5). The derived ozone trend is shown in Fig. 3. A maximum change of $\sim 7\%$ occurs at ~ 46 km. This change compares reasonably well with a two-dimensional model calculation, for the same period and latitude range, which includes anthropogenically increasing halogens as well as solar cycle variations.

The two-dimensional model has a latitude range from 85° S to 85° N and an altitude extent from the ground to ~ 90 km (refs 10, 11). The data shown here cover the same range as the observations. Anthropogenically increasing halogen source gases cause upper stratospheric Cl_y (Cl , ClO , HOCl , HCl and ClONO_2) to increase from ~ 1.8 to ~ 2.7 parts per 10^9 (p.p.b.v.) between 1980 and 1989, whereas the ultraviolet solar-cycle variation is responsible for only modest changes in the stratosphere, as 1980 and 1989 are both near solar-cycle maxima. The difference in solar activity between the two years is less than 1% at wavelengths longer than 200 nm (ref. 12). Observations of 27-day solar activity has shown that a 1% solar ultraviolet variation would result in a 1–2% change in 3 mbar ozone¹³. Therefore any differences in solar activity in 1980 and 1989 would have negligible effects on ozone.

Although on the whole the model and observations agree within errors, the model predicts largest ozone decreases near 2.5 mbar whereas the measurements indicate largest decreases near 1 mbar, and the predicted ozone depletions at all altitudes are consistently more than those measured. The model uses a fixed monthly vertical temperature distribution for the 1980 and

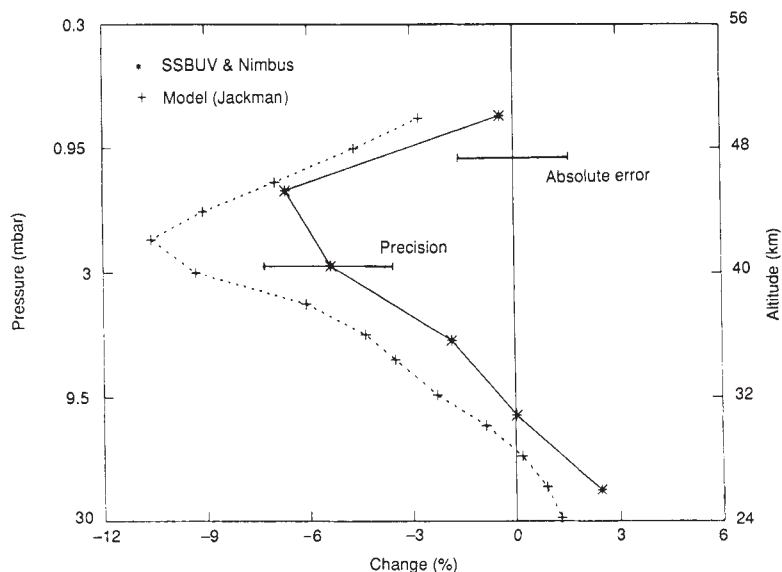


FIG. 3 Ozone trend for October from 1980 to 1989 derived from SSBUV corrected Nimbus-7 data, compared with a two-dimensional photochemical model prediction. Error bars denote systematic and random errors separately.

1989 calculations; it therefore omits the feedback (cooling) on temperature, which should result from ozone depletion and which should mitigate the rate of photochemical destruction of ozone. Although all known photochemical reactions believed to be important in the upper stratosphere are included in the model, it is possible that unknown reactions have been overlooked.

For the observations, we assumed that there were no initial calibration biases between the Nimbus-7 SBUV and SSBUV and therefore that both instruments had the same sensitivity to ozone. The SSBUV calibration procedures and standards have been modelled after the procedures used for the Nimbus-7 SBUV¹⁴. Absolute errors in calibration are systematic and involve the radiometric standards and their application to the instrument calibration. Another component to the absolute error is the calibration stability of SSBUV during flight; but the calibration for the flight is known to better than 1% (ref. 15). The combined systematic accuracy of the two instrument calibrations is bounded by the absolute error bar ($\pm 2\%$, 1σ) shown in Fig. 3. The retrieval algorithm for the two instruments is identical and therefore cancels in our comparison. The second source of errors is random and is bounded by the precision error bar (1σ) in Fig. 3. The random errors are the root sum square of the standard errors of the zonal average ozone amounts, the ozone average over 30°S to 30°N , and the average of the correction factor over latitude.

How do these initial results compare with other observations? The amount of ozone in the layer between 35 and 50 km is $\sim 15\%$ of the total column amount. The 5% change observed in this layer would be undetectable in column ozone trends observed by satellites³. Ground-based Umkehr observations of trends in this altitude layer are consistent with the results presented here but remain tentative⁵. SAGE results¹⁶ show trends similar to those shown here, although their period is 1979 to 1990. The SAGE data for the latitude, altitude and time period shown here indicate (refer to Fig. 3): 1% ozone depletion at ~ 50 km, a maximum depletion of $\sim 4\%$ at 43 km, zero change at ~ 34 km, and an ozone increase of 3% at 30 km. The 1σ errors for SAGE are $\sim 2\%$, roughly comparable to ours. Although the results overlap within the errors, the SAGE trend seems to be systematically lower than our observations and model calculations.

SSBUV data will have more effect on our ability to detect ozone trends in the upper stratosphere after several flights have been made and data merged with the NOAA satellite data. The upper stratospheric ozone trends discussed here make a small contribution to the total ozone trend³; most of the total ozone decrease results from ozone being depleted in the lower stratosphere^{2,16}. Nevertheless ozone changes in the upper levels will have a considerable impact on the radiative properties of the atmosphere, and this effect will then feed back to the temperature distributions and dynamics (winds and mixing) of the stratosphere. $\square$

Model of anomalous transport with backreacting localized perturbations

M. A. Dubois, P. Beaufumé & B. Fromont

DRFC, CEN Cadarache, F-13108 St Paul-lez-Durance, France
I.E.S.L., Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

MANY problems involving transport in physics, astrophysics and biology can be described in terms of a discrepancy between the experimentally observed flux of a quantity U and that predicted by elementary diffusion theory: this is known as anomalous transport. In such cases, there is often evidence for discrete structures or processes, spatially localized at fixed sites within the system, which can enhance the transport by giving rise to intermittent percolating pathways across the system; examples include Kelvin-Helmholtz vortices in astrophysical accretion disks, island structures in plasma magnetic turbulence and enzymatic processes at the interfaces of cells. Such structures may themselves depend on U . To study the nonlinear elementary processes operating in these systems, we introduce a general model comprising localized structures coupled by two sets of variables: the transported quantity $\{U_i\}$, and a measure $\{\delta_i\}$ of the influence of each structure on the diffusion of U . The model is related to, but distinct from, cellular automata, which use only one set of coupling variables. We study the evolution of this model using anomalous transport in a turbulent magnetic plasma as an example. We believe that this model could provide a basis for studying the behaviour of a wide range of nonlinear systems.

We consider a one-dimensional space (a slab or cylinder) with n sites, represented by $\{z_i\}_{i \in [1, n]}$. The variables U and δ are real functions of space and time, which are both taken to be discrete. The active processes (nicknamed 'beasts' until specified) are assumed to be located on these sites along a direction z , which is also the direction along which the quantity U is transported (Fig. 1).

We have two sets of time-dependent real variables: $\{U_i(t)\}_{i \in [1, n]}$ forms a discrete profile of U along the z direction; $\{\delta_i(t)\}_{i \in [1, n]}$ characterizes the state of every 'beast' i (that is, its local action on the diffusion of U); the functions δ_i are graphically displayed as the width of the 'beasts' (see Fig. 1).

The evolution of U is given by the classic diffusion equation $\partial U / \partial t = -\nabla \cdot \phi$. Discretization yields the following set of equations:

$$U_i(t+1) - U_i(t) = (-dt/dz)(\phi_{i+1}(t) - \phi_i(t))$$

with $\phi_i(t) = -\chi_i(t)\nabla U_i(t)$ (Fourier's law)

$$(\delta_i(t+1) - \delta_i(t))/dt = \gamma_i(t) = \Gamma(U_i, \delta_j)_{j \in [0, n]}$$

$$\chi_i = K(U_i, \delta_j)_{j \in [0, n]}$$

where ϕ_i is the flux and χ_i is the diffusion coefficient between site i and site $i+1$. The model is completely described if we prescribe the beasts' evolution law Γ , the diffusion coefficient law K , and the initial and boundary conditions, among which is the source term ϕ_{in} , of entering flux.

Such a model can be used as a test bench to see whether a type of localized physical phenomenon can be held responsible for the global behaviour of a system.

We now address specifically the problem of anomalous transport linked with magnetic turbulence, as may occur in tokamaks and in some astrophysical plasmas. We assume the magnetic structure to be made of nested surfaces, each one being labelled by its helicity which is the ratio of the longitudinal to the azimuthal magnetic field on the surface. In this structure rational surfaces (those with a rational helicity) are intermingled in a bulk of irrational ones. On rational surfaces, also called resonant

Received 19 December 1991; accepted 18 May 1992.

1. Molina, L. T., Tso, T. L., Molina, L. T. & Wang, F. C. Y. *Science* **238**, 1253-1257 (1987).
2. Anderson, J. G. et al. *J. geophys. Res.* **94**, 11480-11520 (1989).
3. Stolarski, R. S., Bloomfield, P., McPeters, R. D. & Herman, J. R. *Geophys. Res. Lett.* **18**, 1015-1018 (1991).
4. *Scientific Assessment of Stratospheric Ozone: 1989*, WMO Rep. 20, 163-219 (1990).
5. *Report of the International Ozone Trends Panel: 1988*, WMO Rep. 18, 387-440 (1990).
6. Hilsenrath, E., Williams, D. E. & Frederick, J. E. *Proc. SPIE: Recent Advances Sensors, Radiometry, and Data Processing for Remote Sensing* **924**, 215-222 (1988).
7. *National Plan for Stratospheric Monitoring and Early Detection of Change, 1988-1997*, FCM-P17-1989 (U.S. Department of Commerce/NOAA, 1989).
8. Frederick, J. E. & Niu, X. Hilsenrath, E. *J. atm. ocean. Technol.* **7**, 734-740 (1990).
9. Lean, J. *Rev. Geophys.* **29**, 505-535 (1991).
10. Douglass, A. R., Jackman, C. H. & Stolarski, R. S. *J. geophys. Res.* **94**, 9862-9872 (1989).
11. Jackman, C. H., Douglass, A. R., Rood, R. B., McPeters, R. D. & Meade, P. E. *J. geophys. Res.* **95**, 7417-7428 (1990).
12. DeLand, M. T. & Cebula, R. P. *Proc. Workshop Solar Electromagnetic Radiation Study for Solar Cycle 22* (ed. Donnelly, R. F.) (NOAA ERL, in the Press).
13. Hood, L. L. & Jirlikow, J. L. *J. geophys. Res.* **96**, 7565-7577 (1991).
14. Cebula, R. P., Hilsenrath, E. & Guenther, B. *Proc. SPIE: Optical Radiation Measurements* **1109**, 205-218 (1989).
15. Hilsenrath, E., Cebula, R. P., Hynes, S. & Caffrey, R. *Metrologia* **22**, 301-304 (1991).
16. McCormick, M. P., Veiga, R. E. & Chu, W. P. *Geophys. Res. Lett.* **19**, 269-272 (1992).

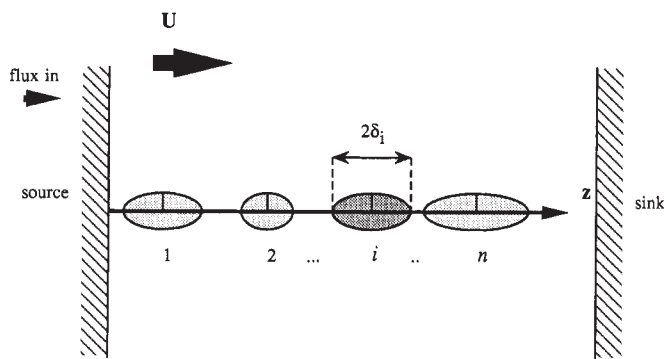


FIG. 1 Geometry of the model; the quantity U is transported across the n sites of the 'beasts'.

surfaces, plasma instabilities create magnetic islands (see ref. 3). Theory predicts that for a rising level of instability, more and more surfaces break down, and that for a critical size of island (corresponding roughly to spatial overlap of the islands) large-scale stochasticity will set in, considerably enhancing transport across the surfaces, whereas it is otherwise limited to collisional diffusion. Let us consider U as a density of monokinetic electrons and the 'beasts' as magnetic islands, located on rational magnetic surfaces. In a first approximation these surfaces are taken to be equidistant. For the local diffusion coefficient χ_i between site i and site $i+1$, we use an analytical diffusion

coefficient in a semi-broken magnetic topology⁴:

$$\chi = \chi_1 \chi_2 / (\chi_1 \min(1, k/k_c)^{1/2} + \chi_2 (1 - \min(1, k/k_c)^{1/2}))$$

where $\chi_1 = \chi_{\text{collisions}}$ and $\chi_2 = (k^2/4)(1 + \beta k^{-\gamma})^{-1}$ with $\beta = 32$, $\gamma = 3$ and $k = ((\delta_i + \delta_{i+1})/(z_{i+1} - z_i))^2$. In this case δ_i and z_i are the islands' half-widths and positions, respectively, and k_c is the value of k for which large-scale stochasticity sets in.

For the δ_i evolution law we take, as an example, the critical density gradient law, as suggested in ref. 5

$$\gamma_i = \Gamma(U_j, \delta_j) = c(|\nabla U_i / G_0| - 1)$$

where G_0 is a critical gradient and c is the beasts' reactivity.

We prescribe initial conditions for the system

$$(U_i(0), \delta_i(0))_{i \in [1, n]}$$

and the following boundary conditions. (1) A constant entering flux, ϕ_{in} , of the quantity U at site 1 is provided by a source and acts as an exterior forcing parameter. (2) At the last site n , a totally absorbing sink, $U_n(t) = \text{constant} = 0$.

Among the parameters involved in the model ($n, G_0, c, \phi_{\text{in}}$), we find that the two leading control parameters are c and ϕ_{in} (we consider the system's asymptotic behaviour for which the U profile has reached a slope close to G_0 ; the value of G_0 acts mainly on the duration of the transition phase). From now on we set n and G_0 constant. In our specific case, δ is the island extension along z . We plot, at their fixed location in z , the island widths as functions of time: overlap in the graphs corresponds to real spatial overlap and therefore to strong diffusion (Fig. 2).

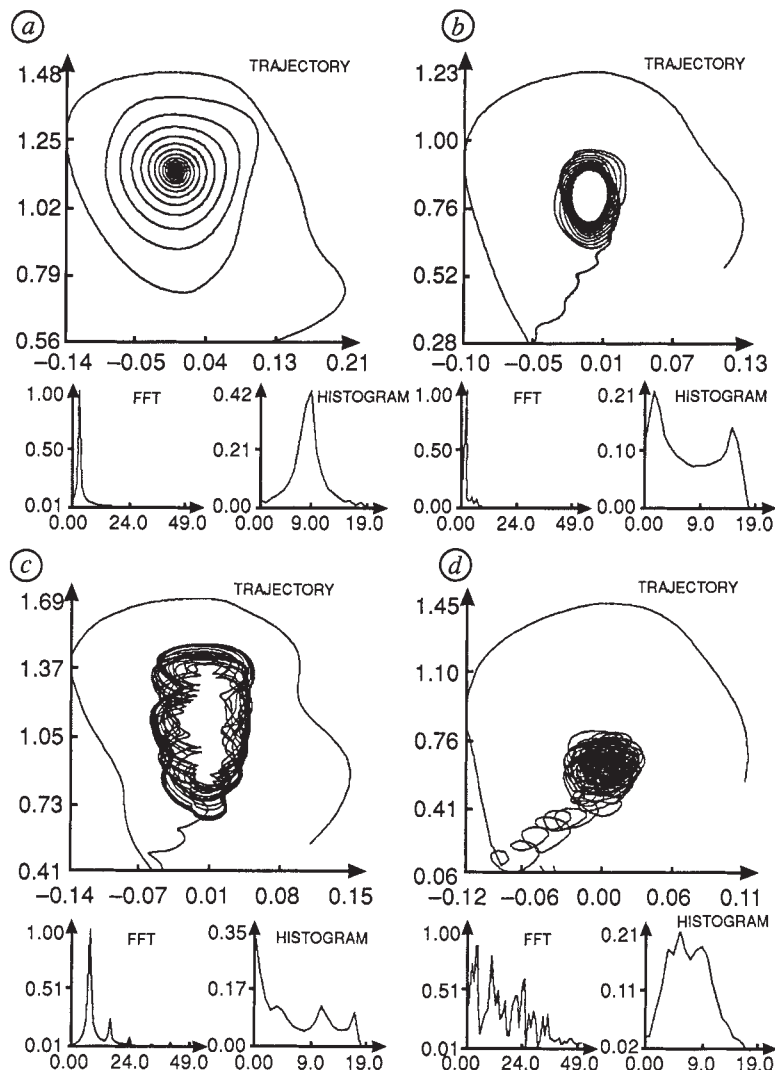


FIG. 3 Two-dimensional trajectory in the integral projection space: the main plot shows W (the integral of U conveniently normalized) on the horizontal axis and T (global measure of turbulence) on the vertical axis. Smaller plots show the Fourier transform and histogram of $W^2 + T^2 = f(t)$. a, Attracting point (regime 1): $c = 2$, $\phi_{\text{in}} = 2$. b, Limit cycle (regime 2): $c = 5$, $\phi_{\text{in}} = 1$. c, Attracting torus (regime 3): $c = 18$, $\phi_{\text{in}} = 4$. d, Chaotic attractor (regime 4): $c = 16$, $\phi_{\text{in}} = 0.5$.

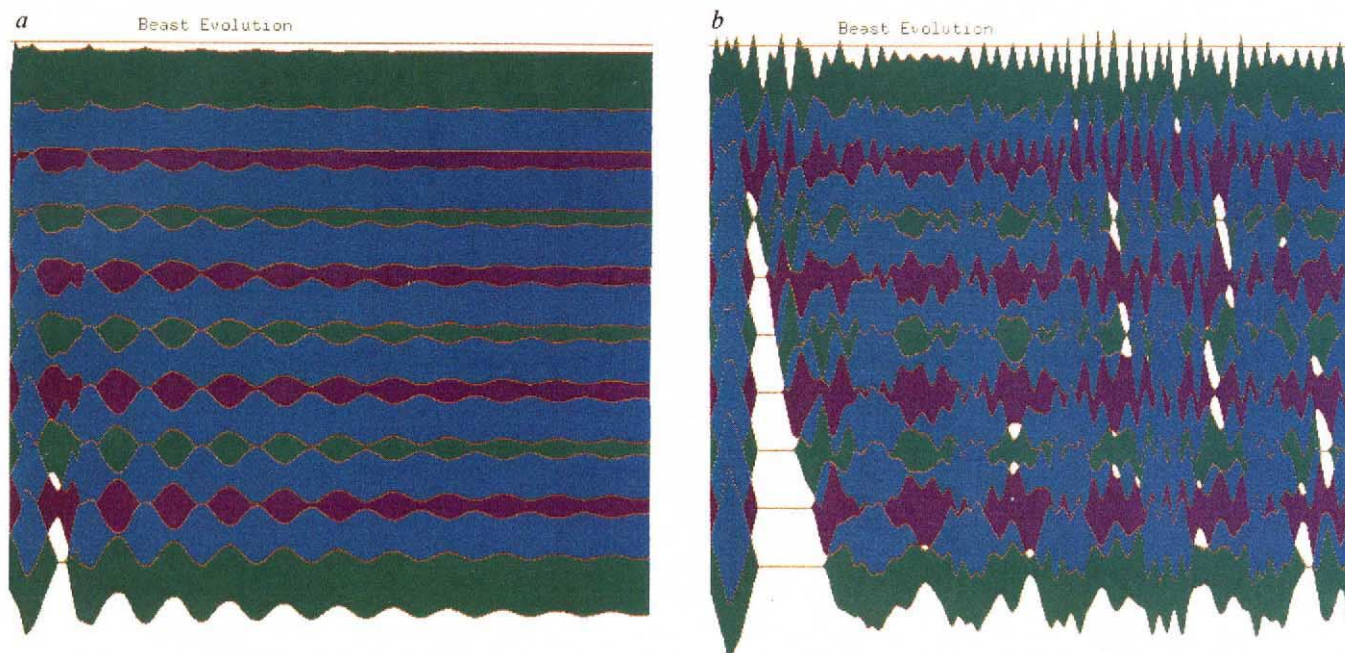


FIG. 2 Evolution of the 'beasts' (distributed on the vertical axis) with time (horizontal axis) for 500 iterations. The widths of the islands are alternatively represented in green and purple, and the spatial overlap between neighbour-

ing islands is shown in blue. For all the cases displayed, we took $n=10$, $G_0=0.9$, $\delta_i(0)=0.25$, $U(0)=10-z_i$, $dt=0.01$. *a*, Attracting point: $c=2$, $\phi_{in}=2$. *b*, Attracting torus: $c=18$, $\phi_{in}=4$.

We call the integral map a two-dimensional projection of the system's trajectory in the following space: horizontal axis, $W \propto \sum U_i$ (integral of the transported quantity); vertical axis, $T = \sum \delta_i$ (global measure of turbulence).

The model has several asymptotic regimes (obtained after a transition phase): (1) attracting fixed point: the trajectory is attracted by a single point; (2) limit cycle: the trajectory tends to or oscillates around a limit cycle; (3) attracting torus: the trajectory consists of the superposition of several modes; (4) chaotic or 'strange' attractor: the trajectory seems to wander randomly in a bounded area.

We show the islands' evolution for regimes 1 and 3 in Fig. 2, and the trajectory in the integral map for the four regimes (Fig. 3).

The temporal evolution of the outgoing flux, ϕ_{out} (the flux of U leaving the last site), closely follows the beasts' activity. It may be very regular in some cases: for regimes of type 1 it tends towards a constant value and for regimes of type 2, it is a nearly perfect sinusoid; but it may also be highly variable: jets of U seem to be ejected from the system in bursts, mainly in regimes of type 4. This behaviour illustrates how the last site acts as a barrier as well as indicating the rapid and marked changes of the gradient of the quantity U , ∇U . (This type of behaviour of the outgoing flux has suggested an application to accretion disks in compact objects; J. Kuijpers and M.A.D., work in progress.)

As we are dealing with a high-dimensional dissipative system showing many unpredictable regimes, an analytical study is hardly feasible, but a systematic exploration of the system's regimes in a large parameter space may lead to important results. We use an intelligent numerical tool dedicated to this aim, 'Explorer'⁶, which is a neural-network-based learning system capable of performing a fast exhaustive study of a given parameter domain. Explorer was taught to classify the various regimes using a Fourier transform and an histogram of $W^2 + T^2 = f(t)$ (see Fig. 3). The fast Fourier transform provides key features to discriminate between some regimes, and the histogram allows further discrimination when needed. The map in Fig. 4 shows Explorer's results in a parameter domain. This chart allows one to locate precisely each regime in the parameter domain. The boundaries between the various zones are far from

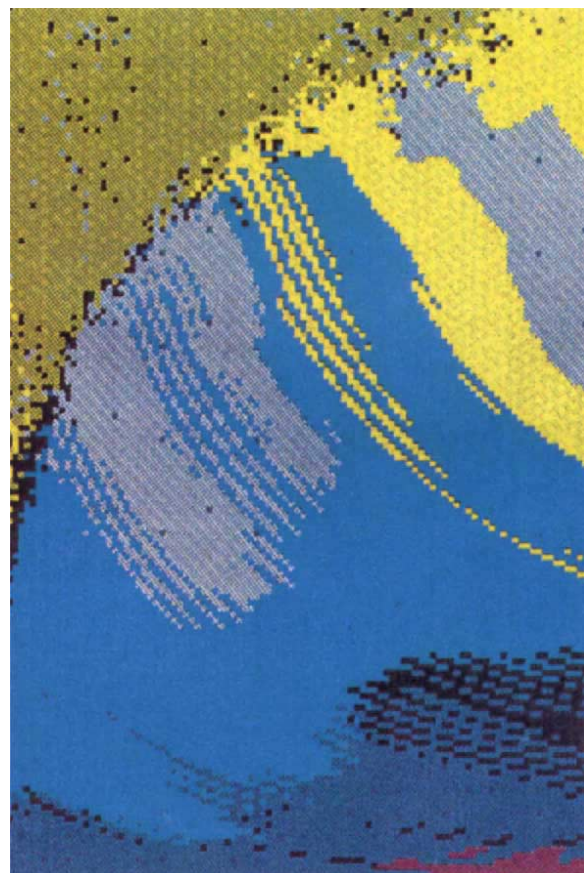


FIG. 4 Exploration of the parameter space (c , ϕ_{in}) using Explorer; the vertical axis is $c \in [1, 20]$, and the horizontal axis is $\phi_{in} \in [0.3, 3]$. Each colour corresponds to one of the regimes or subregimes described in the text, and black dots indicate Explorer's inability to identify with certainty the system's behaviour. Pink represents an 'attracting fixed point'; dark blue, 'spiralling'; light blue, 'limit cycle'; grey, 'perturbed limit cycle'; yellow, 'torus attractor'; and green, 'chaotic attractor'.

being straight lines, reflecting the nonlinear character of the system.

The rich variety of behaviour shown by this class of systems would justify a dedicated study. We are more interested in its value as a tool to study intermittent percolation processes in different physical systems where transport is not understood: for each problem, the choice of the physical meaning of U and δ , growing rate for the beasts, diffusion coefficient for U and the proper geometry will allow us to test possible mechanisms in a fully nonlinear regime. Moreover, comparison of the regime charts of very different physical systems might enlighten us as to the role played by each parameter. In our example, one clearly sees that as c , which characterizes the instability's growth rate, increases, the system's behaviour follows a route to chaos (attracting point, cycle, torus and eventually chaotic attractor); on the other hand, the entering flux ϕ_{in} (here, the total heating power provided to the plasma) forces the system from chaos into order: as the value of ϕ_{in} increases, the 'level of chaoticity' decreases.

The present behaviour can be intuitively understood from the strong correlation of the diffusion coefficient χ with the beasts' amplitude δ , and the rather weak dependence of the growth rate Γ on $(|\nabla U/G_0| - 1)$. It is likely that such charts are not always alike: other systems may have more complex regimes, for instance a trajectory in the integral map with two distinct attracting points. The integral maps themselves are powerful diagnostics of global behaviour. A measure of the percolation efficiency (or conversely of the confinement efficiency) is given by W/ϕ_{in} : in a fusion plasma, this would be analogous to the energy confinement time τ_E .

Our examples are idealized. Taking into account the real locations of resonant surfaces for realistic plasma profiles and real plasma instabilities will lead to predictions directly comparable to tokamak experiments, and we believe that the model will be useful for many other problems in physics. $\square$

Received 23 December 1991; accepted 11 May 1992.

1. Chaté, H. & Manneville, P. *Prog. theor. Phys.* **87**, 1–60 (1992).
2. Wolfram, S. *Physica D* **10** 1–35 (1984).
3. White, R. B. *Basic Plasma Physics I* (eds Galeev, A. A. & Sudan, R. N.) 611–676 (North-Holland, Amsterdam, 1983).
4. Beaufumé, P., Dubois, M. A. & Mohamed Berkadda, M. S. *Phys. Lett. A* **147**, 87–91 (1990).
5. Rebut, P. H., Brusatti, M., Hugon, M. & Lallia, P. P. *Proc. 11th Int. Conf. Plasma Phys. Controlled Nuclear Fusion Res.*, 187 (IAEA, Vienna, 1987).
6. Fromont, B. & Beaufumé, P. *MIT Res. Rep.* 91–15; *Computers Phys.* (submitted).

Origin of ferroelectricity in perovskite oxides

Ronald E. Cohen

Geophysical Laboratory, Carnegie Institution of Washington,
5251 Broad Branch Road NW, Washington DC 20015, USA

FERROELECTRIC materials are characterized by a switchable macroscopic polarization. Most technologically important ferroelectrics are oxides with a perovskite structure. The origin of their ferroelectric behaviour is unclear, however, and there is incomplete understanding of why similar, but chemically different, perovskites should display very different ferroelectric behaviour. The great sensitivity of ferroelectrics to chemistry, defects, electrical boundary conditions and pressure arises from a delicate balance between long-range Coulomb forces (which favour the ferroelectric state) and short-range repulsions (which favour the nonpolar cubic structure). To model the transition accurately, total-energy techniques are required which incorporate the effects of charge distortion and covalency. Here I report results of electronic-structure calculations on two classic examples of ferroelectric perovskites, BaTiO_3 and PbTiO_3 , and demonstrate that

hybridization between the titanium 3d states and the oxygen 2p states is essential for ferroelectricity. The different ferroelectric phase behaviour of the two materials is also clear: in PbTiO_3 , the lead and oxygen states hybridize, leading to a large strain that stabilizes the tetragonal phase, whereas in BaTiO_3 the interaction between barium and oxygen is completely ionic, favouring a rhombohedral structure.

Ferroelectric perovskites provide well known examples of displacive transitions¹. The oxides BaTiO_3 and PbTiO_3 have similar cohesive properties such as unit-cell volume (64.2 and 63.2 Å³, respectively), but different ferroelectric behaviour which has never been understood. Both are paraelectric (non-polar) at high temperatures and have the cubic perovskite structure which is simple cubic (space group $\text{Pm}\bar{3}\text{m}$) with 'A cations' (e.g. Pb and Ba) in the large eightfold coordinated site at (0, 0, 0), 'B cations' (e.g. Ti) in the octahedrally coordinated site at (0.5, 0.5, 0.5), and oxygens at the equipoint (0.5, 0.5, 0). BaTiO_3 has three ferroelectric phase transitions, cubic to tetragonal (393 K), tetragonal to orthorhombic (278 K) and orthorhombic to rhombohedral (183 K), whereas PbTiO_3 has only one, cubic to tetragonal at 766 K. The ferroelectric distortions involve small displacements of the cations relative to the anions, leading to a net dipole moment per unit volume, or polarization. The displacements are similar in the tetragonal phases for the two materials, although larger in PbTiO_3 (refs 2, 3).

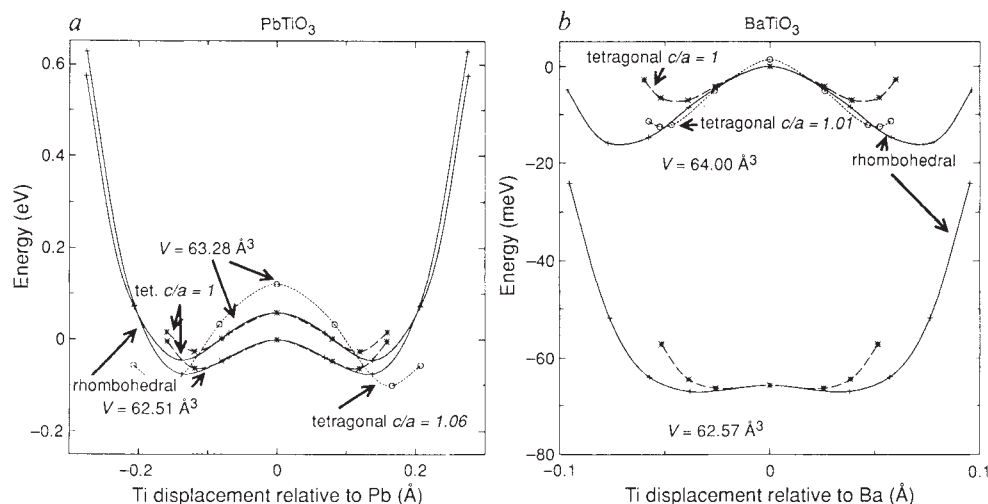
Both materials are also ferroelastic: tetragonal BaTiO_3 has a 1% c/a strain, whereas tetragonal PbTiO_3 shows a large (6%) strain. Typical ferroelastics show strains of 0.5–3% (ref. 4). The relative contributions of the elastic energy, long-range coulomb interactions and local bonding effects to the phase transitions are not known. Also unclear is whether the dynamics should be described as order–disorder or displacive near the phase transitions. The ferroelectric phase transitions in perovskites were long considered prototypical displacive transitions, characterized by a zone-centre vibrational mode, the 'soft mode', with a vanishing frequency at the phase transition and an eigenvector similar to the displacements observed in the ferroelectric phases. Recent work, however, focuses on the order–disorder 'eight-site' model⁵, in which the cubic phase consists of random distortions along eight cube diagonals [111], the tetragonal phase consists of displacements along four cube diagonals giving an average structure with a polarization along [100], the orthorhombic phase has two 'sites' occupied and a polarization along [110], and the rhombohedral phase is ordered along [111].

To ascertain whether this model was appropriate for both BaTiO_3 and PbTiO_3 , I used the all-electron, full-potential, linearized augmented plane-wave method⁶ within the local density approximation (LDA) for the exchange and correlation interactions between electrons⁷. The computations make no assumptions about bonding, ionicity, or the shape of the charge density or potential, and are essentially fully converged. For previous calculations on BaTiO_3 that include extensive convergence tests, see ref. 8.

Ideally the entire parameter space for zone centre and zone boundary distortions, including interactions with volume and shear strains, would be mapped from first principles to predict the phase transition temperatures and average dynamical behaviour; this is not technically feasible, so I have used the experimental distortions as a guide, and consider here only rhombohedral and tetragonal zone-centre distortions.

Figure 1 shows the energy against the experimental distortion for BaTiO_3 and PbTiO_3 , with and without tetragonal c/a strain. Without the strain, both BaTiO_3 and PbTiO_3 give the lowest energy for the rhombohedral structure. Energetically the rhombohedral strain is small (it contributes less than 11 K to the energy differences), but the tetragonal strain has a considerable effect in BaTiO_3 , and in PbTiO_3 it stabilizes the tetragonal structure over the rhombohedral phase. In the high-temperature phases, the underlying multiple-well structure of the potential surface still exists. These phases in both BaTiO_3 and PbTiO_3

FIG. 1 Calculated energy as a function of soft-mode distortion in (a) PbTiO_3 and (b) BaTiO_3 . The PbTiO_3 wells are much deeper than for BaTiO_3 . In both BaTiO_3 and PbTiO_3 without strain the rhombohedral phase has lower energy. In PbTiO_3 , however, tetragonal strain stabilizes the tetragonal phase over the rhombohedral phase. In BaTiO_3 strain has a much smaller effect and the rhombohedral phase is still lower in energy. The large volume effect (pressure decreases the well depths) is more obvious in BaTiO_3 because of the smaller energy scale.



are anharmonic, disordered crystals with the atoms seldom in their idealized cubic positions; the ions are mostly displaced along the $[111]$ directions. The tetragonal phase in PbTiO_3 is ordered because of the large stabilizing strain (the strained experimental structure is more than 1,800 K lower in energy than the cubic structure), but in BaTiO_3 the corresponding energy difference is only ~ 150 K, and only the low-temperature rhombohedral phase is expected to be well ordered, as seen by experiment⁵.

Figure 1 also shows the large volume dependence of the soft-mode potential surfaces, consistent with the loss of ferroelectricity in the titanates at high pressures⁹. The ferroelectric to paraelectric transition with pressure is different from the thermally induced transition. Under compression the multiple-well structure disappears, and a continuous transition at zero temperature occurs at the pressure where the well depths vanish; as the low-pressure transition is first-order, there is a tricritical point where the transition changes from first-order to continuous¹⁰. The large sensitivity to volume makes the volume errors in LDA (here 2–4%) unusually important, and the calculated ground state may in some cases be only slightly distorted or even cubic. Other studies indicate that good estimates of vibrational frequencies and elastic constants are obtained using the LDA for the experimental lattice^{11,12}. Better approximations for exchange correlation would reduce the uncertainty.

I found the elements of the dynamical matrix, expanded around the experimental tetragonal structure of PbTiO_3 , by 'frozen phonon' displacements of the atoms, and the matrix was diagonalized giving the eigenmode frequencies within the harmonic approximation. The three computed A_1 frequencies (displacements in the polar direction) at 142, 446 and 663 cm^{-1} and the four E frequencies (transverse displacements) at 75, 190, 250 and 480 cm^{-1} , are similar to the experimental values¹³ of 148, 362, 650, 89, 220, 290 and 508 cm^{-1} , respectively. The discrepancies are probably due to the anharmonic shape of the potential surface. The lowest frequency is stable, indicating the absence of lower-temperature transitions. BaTiO_3 , which does have lower-temperature transitions, is still dynamically unstable in the tetragonal phase as the ferroelectric distortions are small.

The electronic densities of states (Fig. 2) indicate that Ti 3d states are strongly hybridized with the O 2p and that the hybridization is enhanced by the ferroelectric distortion. The sum of the band eigenvalues is one term in the total energy (the sum double-counts the coulomb potential and includes the exchange-correlation potential functional rather than the energy functional)¹⁴, and it clearly decreases with the ferroelectric distortion due to the Ti–O bonding interaction. It is striking to compare the partial densities of states of PbTiO_3 with BaTiO_3 using the experimental PbTiO_3 structure for both, so that differences are due entirely to the replacement of Ba and Pb. Lead 6s and O

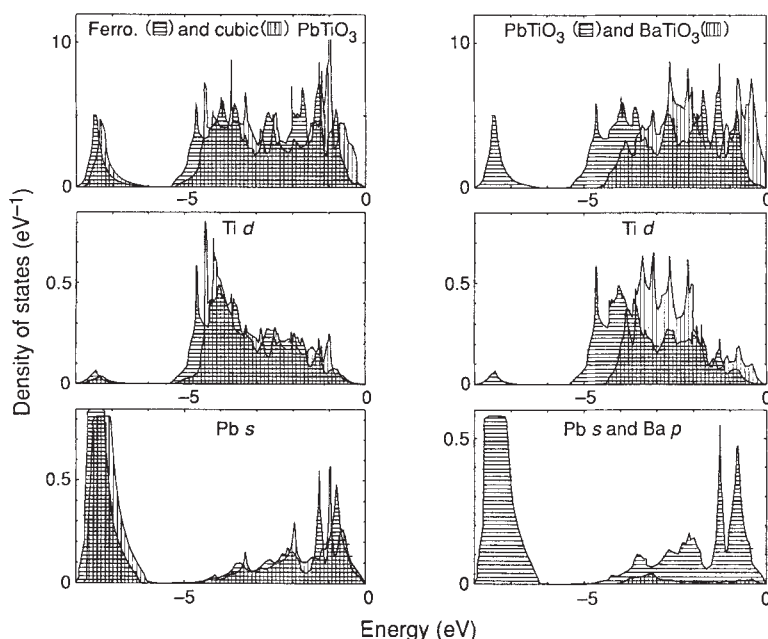


FIG. 2 Electronic density of states against energy relative to the valence-band maxima for (a) ferroelectric PbTiO_3 (horizontal stripes) and cubic PbTiO_3 (vertical stripes), and (b) ferroelectric PbTiO_3 (horizontal) and BaTiO_3 (vertical stripes). The top panels show the total densities of states and the bottom two show the contributions (partial densities of states) from the regions around the Ti for states with d character and around Pb or Ba with s or p character, respectively. Overlap of the Ti 3d and Pb 6s partial densities of states with the primarily oxygen 2p valence bands, which range from 0 to ~ -5.5 eV, indicates considerable hybridization. The ferroelectric distortion also lowers the eigenvalue sum, helping to decrease the total energy. The PbTiO_3 structure was used for BaTiO_3 to obtain the results in the right panels (b) so that the differences are due solely to the replacement of Pb by Ba; whereas the Pb s state is strongly hybridized with the oxygen 2p states, the Ba is fully ionic. The Ti–O interactions are also different in BaTiO_3 and PbTiO_3 even with the same structure, because they are influenced in the latter by the Pb–O interaction.

2p states are strongly hybridized in PbTiO_3 , but Ba 5p does not hybridize with the valence band. The Pb-O bonding interaction, and the smaller ionic radius of Pb^{2+} compared with Ba^{2+} , lead to the larger strain in PbTiO_3 that reduces some Pb-O distances; the Ti-O repulsion prevents the volume shrinking enough to stabilize the cubic phase. There is also an indirect effect on the Ti-O interactions through the Pb-O hybridization; the Ti 3d eigenvalues are lower in PbTiO_3 relative to BaTiO_3 even for the same atomic displacements.

My calculations do not support the idea that the ferroelectric distortion is due to the B cation 'rattling' in the oxygen cage¹⁵ and that the volume is constrained by the A-cation repulsions with oxygen, nor is this idea consistent with radius ratio arguments. The shortest Ti-O distances in the experimental tetragonal structures of BaTiO_3 and PbTiO_3 are 1.86 and 1.78 Å, respectively, significantly smaller than the 2.00 Å expected from the Shannon and Prewitt (S-P) ionic radii¹⁶, whereas the shortest Ba-O and Pb-O distances are 2.79 and 2.53 Å, comparable to the S-P radii (2.82 and 2.69 Å).

The valence charge on the Ti, its polarization, and the dynamical nature of the covalency and polarization, are evident in the charge density (Fig. 3). Dipolar electron density around the Pb (hybrid Pb-O states) develops with increasing ferroelectric dis-

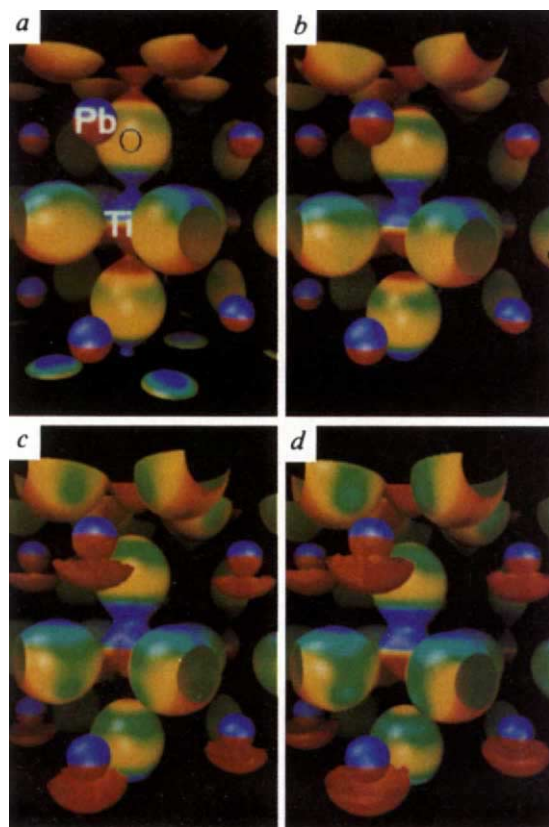


FIG. 3 Valence charge density of PbTiO_3 for increasing ferroelectric distortions (all with $c/a=1.06$). The O and Ti atoms are displaced towards the bottom of the figure. Panels a-d correspond to the four strained tetragonal points in Fig. 1a. Constant density surfaces for 0.26 electrons $\text{\AA}^{-3}$ are shown. The charge density corresponds to the oxygen 2p bands as well as the Pb s band. The most important features are the significant valence charge density on the Ti (Ti^{4+} would have essentially no valence charge) and the clear polarization of the Pb ions. A similar distortion does not occur around the Ba in BaTiO_3 . The colours give the z-component of the electric field at each point on the surface. Near the nuclei the electric field is dominated by the nuclear field, but at greater distances the asymmetry between the red and blue colouring indicates the periodic field that forms in the crystal because of the ferroelectric distortion. A finite crystal could also develop a macroscopic field, depending on the surface boundary conditions.

tortion. In contrast, Ba does not polarize with the ferroelectric distortion. In both materials the Ti strongly distorts and bonds with the closer oxygen. Ti is not fully ionized as a 4+ ion; fits to the self-consistent charge density give $\text{O}^{-1.63}$ and $\text{Ti}^{+2.89}$.

If the Ti-O hybridization is inhibited, the ferroelectric instability disappears and the cubic phase is most stable. This has been shown in two ways. Non-empirical model calculations using the 'potential induced breathing' model, which assume spherical ions and no covalency, give no ferroelectric state for either BaTiO_3 (ref. 17) or PbTiO_3 . Analysis of the results shows that although the long-range Coulomb (Madelung) energy favours a ferroelectric distortion, the short-range repulsions are strong enough to stabilize the cubic phase. Second, in calculations using the linearized augmented plane-wave method if the Ti d energy parameter is raised to reduce the variational freedom, the ferroelectric state is lost, indicating the importance of the Ti d-states for the instability. We can conclude that for ferroelectric perovskites in general, hybridization between the B cation and O is essential to weaken the short-range repulsions and allow the ferroelectric transition. Most ferroelectric oxide perovskites have B cations whose lowest unoccupied states are d-states (Ti^{4+} , Nb^{5+} , Zr^{4+} and so on). This allows for d-hybridization with the O that softens the B-O repulsion and allows the ferroelectric instability at low pressures. The A-cation can modify the ground state and nature of the transition by (1) increasing the ferroelastic strain that couples with the ferroelectric distortions or (2) hybridizing with the valence states, leading indirectly to changes in the B-O interactions. The delicate balance of short-range forces favouring the cubic phase and long-range forces favouring the ferroelectric state makes the transition sensitive to defects that modify the short-range interactions and to carriers (e.g. photoelectrons) that screen the long-range field.

As the temperature of rhombohedral ground-state perovskites such as BaTiO_3 and KNbO_3 is raised, the atoms tend to average over the low-energy [111] distortions, leading to a series of phase transitions. Even in cases with small strain, such as BaTiO_3 , the strain energy strongly affects the phase transitions¹⁸. In tetragonal ground-state perovskites such as PbTiO_3 , the strain is larger and stabilizes an ordered tetragonal phase; at high temperatures, however, the cubic phase PbTiO_3 also has an underlying multiple-well potential surface and should be anharmonic, as is observed. The strain changes the equilibrium phase diagram and should also damp the local fluctuations in the ferroelectric phases. Underlying anharmonic potential surfaces, leading to high-temperature phases with atoms that seldom sit on their equilibrium positions, are not unique to ferroelectrics, and also typify rotational instabilities in perovskites and the tilt instabilities in the cuprate superconductors¹⁹. □

Received 11 March; accepted 8 May 1992.

- Lines, M. E. & Glass, A. M. *Principles and Applications of Ferroelectrics and Related Materials* (Clarendon, Oxford, 1977).
- Shirane, G., Pepinsky, R. & Frazier, B. C. *Acta crystallogr.* **9**, 131-140 (1956).
- Harada, J., Pedersen, T. & Barnea, Z. *Acta crystallogr.* **A26**, 608-612 (1970).
- Salje, E. K. H. *Phase Transitions in Ferroelastic and Co-elastic Crystals* (Cambridge Univ. Press, 1990).
- Comes, R., Lambert, M. & Guinier, A. *Acta crystallogr.* **A26**, 244-254 (1970).
- Wei, S.-H. & Krakauer, H. *Phys. Rev. Lett.* **55**, 1200-1203 (1985).
- Hedin, L. & Lundqvist, B. I. *J. Phys. C*, **4**, 2064-2083 (1971).
- Cohen, R. E. & Krakauer, H. *Phys. Rev.* **B42**, 6416-6423 (1990).
- Decker, D. L. & Zhao, Y. X. *Phys. Rev.* **B39**, 2432-2438 (1989).
- Ramirez, R., Lapeña, M. F. & Gonzalo, J. A. *Phys. Rev.* **B42**, 2604-2606 (1990).
- Cohen, R. E., Pickett, W. E. & Krakauer, H. *Phys. Rev. Lett.* **64**, 2575-2578 (1990).
- Cohen, R. E. in *High Pressure Research: Application to Earth and Planetary Science* (eds Syono, Y. & Manghnani, M. H.) (Terra Scientific, Tokyo, 1992).
- Fontana, M. D., Idrissi, H., Kugel, G. E. & Wojcik, K. J. *Phys. Condens. Matter* **3**, 8695-8705 (1991).
- Kohn, W. & Vashishta, P. in *Theory of the Inhomogeneous Electron Gas* (eds Lundqvist, S. & March, N. H.) (Plenum, New York, 1983).
- Slater, J. C. *Phys. Rev.* **78**, 748-761 (1950).
- Shannon, R. D. & Prewitt, C. T. *Acta crystallogr.* **B25**, 925-946 (1969).
- Boyer, L. L., Mehl, M. J., Flocken, J. W. & Hardy, J. R. *J. appl. Phys.* **24**, Suppl. 24-2, 204-205 (1985).
- Marais, S., Heine, V., Nex, C. & Salje, E. *Phys. Rev. Lett.* **66**, 2480-2483 (1991).
- Pickett, W. E., Cohen, R. E. & Krakauer, H. *Phys. Rev. Lett.* **67**, 228-231 (1991).

ACKNOWLEDGEMENTS. I thank L. L. Boyer, R. J. Hemley, H. Krakauer, W. E. Pickett, and D. Singh for discussions and H. Krakauer, R. M. Hazen and C. T. Prewitt for reading the manuscript. Computations were done on the Cray 2 at the National Center for Supercomputing Applications under the auspices of the NSF. This research was supported by the Office of Naval Research.

Regional trends in mercury distribution across the Great Lakes states, north central USA

Edward A. Nater & David F. Grigal

Soil Science Department, University of Minnesota, St Paul, Minnesota 55108-6028, USA

CONCENTRATIONS of mercury in the environment are increasing as a result of human activities, notably fossil-fuel burning and incineration of municipal wastes. Increasing levels of mercury in aquatic environments and consequently in fish populations are recognized as a public-health problem^{1,2}. Enhanced mercury concentrations in lake sediments relative to pre-industrial values have also been attributed to anthropogenic pollution. It is generally assumed that atmospheric mercury deposition is dominated by global-scale processes, consequently being regionally uniform. Here, to the contrary, we report a significant gradient in concentrations and total amounts of mercury in organic litter and surface mineral soil along a transect of forested sites across the north central United States from northwestern Minnesota to eastern Michigan. This gradient is accompanied by parallel changes in wet sulphate deposition and human activity along the transect, suggesting that the regional variation in mercury content is due to deposition of anthropogenic mercury, most probably in particulate form.

The overall global cycle of Hg deposition is fairly clear, although a detailed understanding of many mechanisms is lacking³. Mercury is released to the atmosphere either by natural processes from soils or sediments, or by anthropogenic activities. Most atmospheric Hg (>95%) is in the elemental form, Hg⁰. Because of its gaseous state, residence times for Hg⁰ are high; current estimates are 0.7–2 yr (ref. 4). A small fraction of Hg⁰ present in the upper atmosphere is oxidized by ultraviolet radiation or by interaction with free radicals or ozone to produce Hg²⁺, which is soluble in water and readily scavenged from the atmosphere by precipitation³.

Until recently this global cycle was believed to dominate Hg deposition; local- or regional-scale processes were assumed to be negligible. But a small portion (<5%) of total atmospheric Hg over continental regions is now known to be associated with particulates^{5,6}. The mode of deposition of a particulate fraction differs strongly from that of Hg⁰, being dominated by dry deposition and precipitation scavenging. Indeed, estimates of deposition velocities suggest that one-third of the total annual input of Hg in central Wisconsin comes from dry deposition of particulates⁷. Estimates of scavenging ratios further suggest that most Hg in precipitation in continental regions is derived from wash-out of particulate Hg (refs 5, 7). Consequently, particulates usually have relatively short atmospheric residence times and their deposition often shows local- or regional-scale spatial patterns.

If the only significant mode of deposition of atmospheric Hg is associated with the global cycle, then Hg distribution in terrestrial systems should be relatively uniform on a regional basis. If Hg deposition is dominated by regional processes, however, geographical patterns in the concentration and amounts of Hg in terrestrial systems should reflect patterns of increasing human activity.

Direct determination of regional differences in deposition of Hg in precipitation is difficult because high temporal variability in the concentration in, and the quantity of, precipitation reduces the accuracy of estimates based on only a few years' data⁸. Additionally, these measurements do not account for dry deposition, which is usually significant for particulates. Analysis of materials that integrate Hg deposition over a long period is an indirect but more accurate method of assessing regional differences

in deposition. Lake sediments yield such records of deposition, but spatial and temporal variability in sediment deposition and variable contributions from terrestrial watersheds limit their usefulness⁹. No significant geographic distribution pattern in Hg deposition has yet been identified in the Great Lakes states by means of either precipitation or sediments.

Soils also accumulate Hg, retaining up to 90% of the Hg currently deposited to terrestrial landscapes^{10,11} as organic matter complexes or on cation exchange sites of clays¹², or by accumulating particulate forms. The use of soil Hg levels to estimate Hg deposition from the atmosphere is not entirely straightforward because variable contributions from geological sources can confound the results and because post-depositional mobility of Hg may alter atmospheric signals.

We hypothesized that Hg in soils would show clear patterns of distribution paralleling trends of increasing industrialization and density of anthropic sources. We tested this hypothesis by sampling upland forest sites along a transect of increasing wet sulphate deposition^{13,14} from northwestern Minnesota to eastern Michigan. For details of field sampling and laboratory sample preparation, see refs 15, 16. Briefly, 155 forest stands were randomly selected to balance sampling among five forest types and five geographical zones across the forested portions of Minnesota, Wisconsin, and Michigan. The five forest types considered were balsam fir (*Abies balsamea* (L.) Mill.), northern hardwoods dominated by sugar maple (*Acer saccharum* Marsh.), jack pine (*Pinus banksiana* Lamb.), red pine (*P. resinosa* Ait.), and aspen (*Populus tremuloides* Michx.). The zones (Fig. 1) were established to correspond to isolines of wet sulphate deposition^{13,14}. Both forest floor (surficial organic litter) and surface mineral soil (0–25 cm depth) were sampled at five locations in each stand. Deeper mineral soil samples (75–100 cm) were collected at three locations in each stand and composited in the field.

Samples of forest floor and surface soil were composited to yield single samples of each type for each stand. All samples were air-dried and digested at 145 °C with sequential additions of concentrated HNO₃, H₂SO₄ and HCl, followed by cold digestion with KMnO₄ and K₂S₂O₈. Mercury was determined by cold vapour atomic absorption spectrophotometry in conjunction with a gold trap¹⁷. Coefficients of variation for duplicates of forest floor and mineral soil samples were 13 and 15%, and recoveries of standard additions averaged 111 and 114%, respectively. One-way analyses of variance¹⁸ were used to determine significant differences between variables by geographical zone. We treated forest type as a block in the analyses¹⁸.

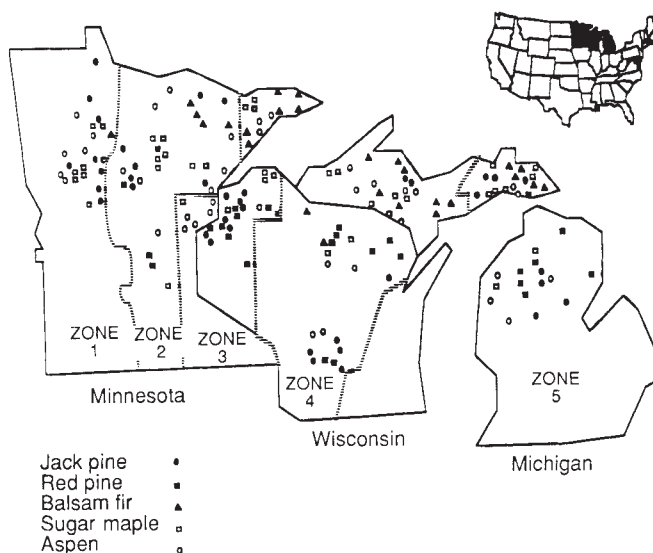


FIG. 1 Sample locations and the five zones established across the Great Lake states.

Separation of means was by Bayes least significant difference¹⁹. Linear regression analyses were by SYSTAT²⁰.

Mercury concentrations in forest floor material generally increased from west (Zone 1) to east (Zone 5) (Fig. 2a). This pattern of increase has also been shown for lead and cadmium in the same samples, but not for base cations nor for phosphorus^{15,16}. Concentrations of Hg in surface soil also increased from west to east, but with a decrease in the most easterly zone (Fig. 2a).

We were concerned about the influence of geological substrate (soil parent material) on the trends in Hg concentration. Because our transect crossed highly variable geological materials (calcareous and non-calcareous glacial tills, andesitic bedrock, shales and alluvium) which were not uniformly distributed²¹, we did not expect a regular trend in Hg from that source. Analysis of 25 random samples from the deepest soil horizon, representing all zones, showed no significant differences in Hg concentrations among zones ($F = 1.134$, 4 and 20 d.f., $P = 0.37$, $MSE = 105.9$), with an overall mean of 13.7 ng g^{-1} , standard error 2.1 ng g^{-1} . Moreover, the relationship between Hg concentrations in surface and subsurface mineral soil samples was poor ($r^2 = 0.163$). In the deep horizon, however, there was a close relationship ($r^2 = 0.887$) between silt-plus-clay content and Hg concentration

$$\text{Hg (ng g}^{-1}\text{)} = 5.12 + 0.523 (\text{silt} + \text{clay \%}) \quad (1)$$

Not surprisingly, native Hg is apparently associated with silt and clay.

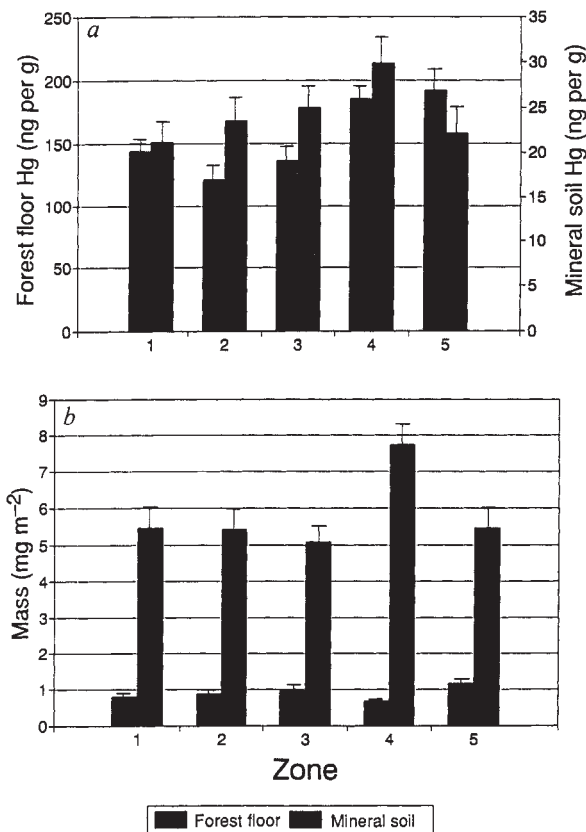


FIG. 2 Mean mercury (Hg) in forest floor and in surface mineral soil (0–25 cm) in five zones along a transect from northwestern Minnesota to eastern Michigan. Means are based on one-way analyses of variance, evaluating significance of differences among zones and blocking on five forest types. Standard errors indicated. a, Differences in concentrations among zones were significant for both forest floor and surface soil. Bayes least significant difference¹⁹ for forest floor concentration was 30 ng g^{-1} , and for surface soil concentration was 7.5 ng g^{-1} . b, Differences in mass among zones were also significant. Bayes least significant differences¹⁹ were 0.23 mg m^{-2} for forest floor mass and 1.5 mg m^{-2} for surface soil mass.

Although concentration is important, the total mass of Hg more closely relates to atmospheric deposition and potential movement of Hg to aquatic systems. The total mass of Hg in forest floor differed significantly among zones ($F = 5.88$, 4 and 109 d.f., $P = 0.0003$, $MSE = 0.178$) (Fig. 2b). The total mass of Hg in surface mineral soil was more than five-fold greater than in forest floor (Fig. 2b) and differences were also significant among zones ($F = 4.83$, 4 and 131 d.f., $P = 0.0011$, $MSE = 8.541$). By this measure, zone 4 had a significantly higher mass than the other zones (Fig. 2b).

Substituting the silt-plus-clay contents of the surface horizons into equation (1) yielded a statistical prediction of the geological contribution or background levels of Hg in those horizons; it also differed by zone ($F = 6.88$, 4 and 145 d.f., $P = 0.0001$, $MSE = 4.87$) (Fig. 3). Subtracting this contribution from the total mass of Hg in the surface horizon yielded the estimated change in Hg mass, which also increased from west to east with significant differences among zones ($F = 3.91$, 4 and 131 d.f., $P = 0.005$, $MSE = 6.563$). Surface soils in the three westerly zones currently contain less Hg than background, indicating that losses of Hg have exceeded inputs. Conversely, the two easternmost zones contain 1.1 and 1.3 mg m^{-2} more Hg than background, indicating enrichment.

When the Hg mass in forest floor is added to the change in Hg mass in the surface soil, all zones show enrichment, and clear west-to-east trends persist, with the highest net change in the easternmost sites (Fig. 3). If this sum represents the increase in Hg mass due to atmospheric inputs (total atmospheric inputs minus leaching and volatilization losses), then the minimum atmospheric input ranges from <0.5 to 2.5 mg m^{-2} (Fig. 3). Easterly zones are not only nearer industrial sources of pollution but also receive more precipitation than westerly zones, ranging from 64 cm annual precipitation in Zone 1 to 81 cm in Zone 5. The observed five-fold differences in net Hg increase, however, cannot be wholly attributable to precipitation, which only varies by a factor of 1.25 across the region.

From current rates of atmospheric deposition in the region, $\sim 15 \text{ } \mu\text{g Hg m}^{-2} \text{ yr}^{-1}$ (refs 3, 22), our estimated net increases of 0.5 – 2.5 mg m^{-2} correspond to ~ 35 – 165 years of deposition. Most of the net increase can be accounted for by post-settlement deposition, with the remainder resulting from pre-industrial deposition, estimated to be $\sim 3.3 \text{ } \mu\text{g Hg m}^{-2} \text{ yr}^{-1}$ (ref. 22). If leaching and volatilization losses are similar across this region, then the observed trends result from increased atmospheric deposition of Hg in the more eastern portions of the region.

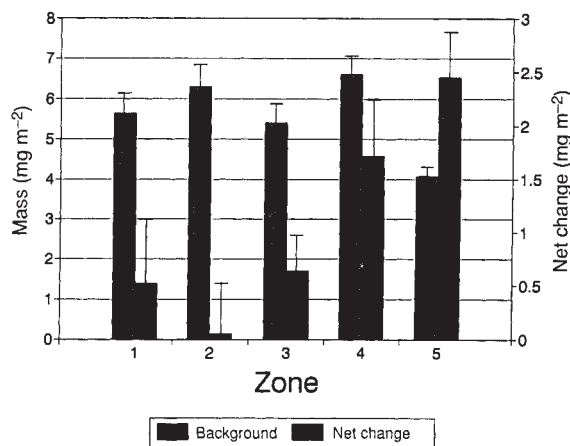


FIG. 3 Predicted geological contributions (labelled 'Background') and estimated net Hg increase resulting from atmospheric deposition in five zones along a transect from northwestern Minnesota to eastern Michigan. Standard errors indicated. Bayes least significant differences¹⁹ were 1.1 mg m^{-2} for background mass and 1.6 mg m^{-2} for net increase.

Lack of knowledge of terrestrial transport and loss processes means that regional differences in post-depositional mobility of atmospheric Hg cannot be excluded as a possible contributor to the observed trends. A gradient in Hg deposition is consistent, however, with west-to-east trends of increasing levels of other atmospheric pollutants in these and associated samples^{15,16,23,24}.

Our data strongly suggest that atmospheric deposition of Hg has a regional component, a result consistent with dry deposition and precipitation scavenging of particulate Hg (refs 5–7). This trend is also consistent with a second possible dry deposition mechanism, stomatal uptake of gaseous Hg⁰ by higher plants²⁵.

If atmospheric Hg⁰ concentrations decrease away from local sources, as suggested by studies in Sweden²⁶, those gradients should produce regional Hg gradients in leaf tissues and hence in leaf litter and soil organic matter. Clear regional patterns of terrestrial Hg distribution have been observed in Norway²⁷ in feather moss (*Hylocomium splendens*) and in Sweden²⁸ in mor layers (compact surficial deposits of organic material) in forested soils. Our results are also supported in the Great Lakes region by a study of fish (walleye, *Stizostedion vitreum vitreum* Mitchell) in lakes with acid neutralizing capacities <300 $\mu\text{eq l}^{-1}$ wherein tissue Hg concentrations increased from west to east²⁹. □

Received 19 December 1991; accepted 19 May 1992.

1. Minnesota Fish Consumption Advisory, Minnesota Department of Health (Minneapolis, 1989).
2. Health Advisory PUBL-1E-019 (89REV), Wisconsin Department of Natural Resources (Madison, 1988).
3. Fitzgerald, W. F. in *The Role of Air-Sea Exchange in Geochemical Cycling* (ed. Buat-Ménard, P.) 363 (Reidel, Boston, 1986).
4. Slemr, F., Seiler, W. & Schuster, G. *J. geophys. Res.* **86**, 1159–1166 (1981).
5. Fitzgerald, W. F., Mason, R. P. & Vandal, G. M. *Water Air Soil Pollut.* **56**, 745–767 (1991).
6. Xiao, Z. F., Munthe, J. & Lindqvist, O. *Water Air Soil Pollut.* **56**, 141–151 (1991).
7. Brosset, C. *Water Air Soil Pollut.* **17**, 37 (1982).
8. Buat-Ménard, P. & Duce, R. A. in *Lead, Mercury, Cadmium, and Arsenic in the Environment* (eds Hutchinson, T. C. & Meema, K. M.) (Wiley, New York, 1987).
9. Evans, R. D. *Arch. Envir. Contam. Toxicol.* **15**, 505–512 (1986).
10. Mierle, G. *Envir. Toxicol. Chem.* **9**, 843–851 (1990).
11. Aastrup, M., Johnson, J., Bringmark, E., Bringmark, I. & Iverfeldt, Å. *Water Air Soil Pollut.* **56**, 155–167 (1991).
12. Andersson, A. in *The Biogeochemistry of Mercury in the Environment* (ed. Nriagu, J. O.) 79–109 (Elsevier/North-Holland, New York, 1979).
13. Verry, E. S. & Harris, A. R. *Water Resour. Res.* **24**, 481–492 (1988).
14. Harris, A. R. & Verry, E. S. in *Hydrological and Hydrogeochemical Mechanisms and Model Approaches to the Acidification of Ecological Systems*, Proc. Int. Hydrological Progr. Workshop (ed. Johansson, I.) 57–65 (Coordinatory Committee for Hydrology, Oslo, 1985).
15. Grigal, D. F. & Ohmann, L. F. *J. envir. Qual.* **18**, 368–373 (1989).

16. Ohmann, L. F. & Grigal, D. F. *USDA Forest Services Res. Bull.* NC-130 (1991).
17. Nater, E. A. & Grigal, D. F. *Report to the Minnesota Pollution Control Agency St Paul, Minnesota* (1992).
18. Snedecor, G. W. & Cochran, W. G. *Statistical Methods*, 6th Edn (Iowa State Univ. Press, 1967).
19. Smith, C. W. *Agron. J.* **70**, 123–127 (1978).
20. Wilkinson, L. *SYSTAT: The System for Statistics* (SYSTAT Inc., Evanston, 1986).
21. Ohmann, L. F., Grigal, D. F. & Brovold, S. *USDA Forest Service Res. Bull.* RB-NC-110 (1989).
22. Brigham, M. thesis, Univ. of Minnesota (1991).
23. David, M. B., Grigal, D. F., Ohmann, L. F. & Gertner, G. Z. *Can. J. Forest Res.* **18**, 1386–1391 (1988).
24. Ohmann, L. F. & Grigal, D. F. *Can. J. Forest Res.* **20**, 508–513 (1990).
25. Lindberg, S. E., Turner, R. R., Meyers, T. P., Taylor, G. E. Jr & Schroeder, W. H. *Water Air Soil Pollut.* **56**, 577–594 (1991).
26. Iverfeldt, Å. *Water Air Soil Pollut.* **56**, 251–265 (1991).
27. Steinnes, E. & Andersson, E. M. *Water Air Soil Pollut.* **56**, 391–404 (1991).
28. Johansson, K., Aastrup, M., Andersson, A., Bringmark, I. & Iverfeldt, Å. *Water Air Soil Pollut.* **56**, 267–281 (1991).
29. Lathrop, R. C., Rasmussen, P. W. & Knauer, D. R. *Water Air Soil Pollut.* **56**, 295–307 (1991).

ACKNOWLEDGEMENTS. We thank L. Kernik, D. Leibl, J. Liddell, J. Lessmeister and S. Brovold for technical assistance, and P. Brezonik and M. Brigham for helpful discussions. This research was funded by the Minnesota Pollution Control Agency under its Acid Deposition Program, by the USDA Forest Service North Central Forest Experiment Station under the National Vegetation Survey Research Program of the National Acid Precipitation Assessment Program and the Northern Stations' Global Change Research Program, and by the Minnesota Agricultural Experiment Station.

Petrogenesis of an augite-bearing melt rock in the Chicxulub structure and its relationship to K/T impact spherules in Haiti

David A. Kring & William V. Boynton

Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona 85721, USA

GEOPHYSICAL anomalies on the Yucatán peninsula define a buried circular structure with an approximate diameter of 180 km (refs 1–3). These anomalies, along with stratigraphic and petrological data, including evidence for shock metamorphism, have been used to interpret the structure as an impact crater⁴. This structure, known as Chicxulub, is particularly interesting because it formed at or near the end of the Cretaceous period, in the geographical region where an impact is believed to have occurred, in large part because of a thick ejecta deposit found on Haiti⁵. Glassy tektite-like relics in this deposit^{6–9} are unusually calcic (up to 31 wt % CaO; ref. 7), providing a further circumstantial link with the Chicxulub structure, which penetrates a carbonate and evaporite sequence^{7,8}. Here we strengthen this link by showing that a simple chemical relationship exists between the glasses and an augite-bearing melt rock found within the Chicxulub structure, and argue that the composition of this melt rock could not easily have been produced by volcanic processes.

The melt rock (Y6N17)⁴ occurs in a 380-m-thick unit near the bottom of a ~1-km-deep borehole drilled into the Chicxulub structure by Petróleos Mexicanos. The melt unit overlies at least 6 to 8 m of anhydrite and lies below a unit consisting in part of polymict breccia, which in turn lies beneath a normal carbonate and evaporite platform sequence. Whereas the melt rock seems to be limited to the interior of the structure, the interior polymict

breccia may be stratigraphically contemporaneous with a brecciated unit outside the structure which was penetrated by a borehole 50 km southeast of the structure's rim⁴.

The melt rock is dominated by a microcrystalline (5 to 10 μm) groundmass consisting of augite, alkali feldspar, plagioclase feldspar, and minor amounts of magnetite, ulvöspinel and apatite. The melt lacks any coherent flow structure, although augite crystals are sometimes aligned in localized regions. Xenocrysts and xenoliths of quartz, up to 1.1 mm in length, occupy 2% of the sample. These clasts are surrounded by coronas of medium-grained augite and alkali feldspar, with occasional micrometre-sized silica relics, which apparently formed in SiO₂-enriched zones where dissolving silica was incompletely mixed with the bulk melt. These coronas, which occupy 13% of the sample, are similar to augite and alkali feldspar coronas surrounding quartz crystals in the Manicouagan impact melt sheet¹⁰ and they resemble the augite plus SiO₂, K₂O-rich glass coronas seen at the Mistastin impact crater^{11–15}. The melt also contains pods of microcrystalline alkali and plagioclase feldspar, occupying 11% of the sample. These feldspar pods, one of which is sieve-textured, seem to represent incompletely assimilated feldspar clasts^{12,15,16} or shock-induced feldspar melts that were incompletely mixed with the bulk melt^{12,14}, such as those seen at other impact structures. The melt rock also contains anhydrite, and although a portion of the anhydrite may represent relic clasts, some of it occurs in veins and seems to be secondary. These veins of anhydrite, in addition to those of quartz, are an expected consequence of hydrothermal processes in large impact craters^{14,17–19}, particularly in one produced on the Yucatán peninsula, which contains a sequence of evaporite deposits and which would have been subsequently covered with sea water.

Although the coronas, feldspar pods and anhydrite, in addition to a lack of plagioclase phenocrysts and orthopyroxene⁴, provide circumstantial evidence that the melt rock was produced by an impact, they are not diagnostic because it is difficult to discriminate these from similar features in volcanic rocks^{20–24}. The principal reason for interpreting the melt rock as an impact

TABLE 1 Average composition of the groundmass in Y6N17 and other relevant compositions

wt% oxide	Pyroxene (24)*	Feldspar (11)*	Feldspar (5)*	Average Y6N17 groundmass†	Bulk Y6N17‡	Average G6 Haitian glass ^{4,8}	Model mixture§
SiO ₂	50.2±0.7	57.8±1.2	70.1±0.2	58.2±1.4	64.0	63.2	63.2
TiO ₂	0.7±0.1	0.1±0.0	0.0±0.0	0.2±0.0	0.4	0.7	0.8
Al ₂ O ₃	2.8±0.7	25.9±0.6	21.0±0.1	18.3±0.5	12.8	15.5	15.2
CaO	23.5±0.4	9.2±0.8	0.2±0.0	10.3±0.3	7.5	7.9	7.8
MgO	12.3±0.8	0.1±0.0	0.0±0.0	3.2±0.1	3.2	2.6	2.6
FeO	10.3±1.1	1.7±0.2	0.2±0.0	3.5±0.1	—	5.4	5.3
Fe ₂ O ₃	—	—	—	—	4.6¶	—	—
Cr ₂ O ₃	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	—	0.0	—
MnO	0.4±0.1	0.0±0.0	0.0±0.0	0.1±0.0	0.1	0.2	0.1
NiO	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	—	0.0	—
Na ₂ O	0.5±0.1	5.9±0.4	10.6±0.2	5.7±0.2	4.1	2.7	2.4
K ₂ O	0.0±0.0	0.6±0.1	0.1±0.0	0.3±0.0	1.9	1.7	2.5
P ₂ O ₅	0.5±0.1	0.2±0.0	0.0±0.0	0.2±0.0	0.1	0.3	0.2
BaO	—	—	—	—	0.1	—	—
LOI#	—	—	—	—	1.4	—	—
Total	101.2	101.4	102.2	100.0	100.2	100.2	100.1

* Number of electron microprobe analyses used to calculate an average composition and its standard deviation.

† Determined by modal reconstruction of electron microprobe analyses. Two thousand points were used to identify the relative percentages of phases.

‡ X-ray fluorescence analysis; corrected for anhydrite.

§ 12.2% Y6N17 groundmass +71.7% SDC-1 mica schist³⁶ +3.9% sandstone +1.8% dolomite +10.4% limestone or anhydrite.

|| Total iron calculated as FeO.

¶ Total iron calculated as Fe₂O₃.

Lost on ignition.

product is its association with an overlying polymict breccia which contains unambiguous shocked quartz⁴.

One method commonly used^{12,14,19,25,26} to demonstrate that a melt rock had an impact origin is to show that it is a chemical mixture of the suspected target lithologies produced by wholesale melting of them. This method is impossible to apply in the case of Chicxulub because an impact crater this size would have excavated material to a depth of 13–15 km (ref. 4) and displaced material to a depth of 39–44 km (ref. 27), where samples of the crust are not available.

An alternative method, used here, is to demonstrate that the Chicxulub melt rock could not have formed as a consequence of common volcanic processes, in particular by showing that the liquid represented by the augite-bearing groundmass does not lie on chemical trends produced by igneous fractionation processes (with or without assimilation). To illustrate the chemical trends typical of relevant igneous suites (those of the calc-alkaline and tholeiitic series), we use a pseudoternary projection scheme^{28–30} (Fig. 1a). In such a diagram parental magmas for both the calc-alkaline and tholeiitic series lie in the olivine- and plagioclase-saturated field. When these melts evolve they move towards the liquid and silica-rich eutectoids, even when they assimilate crustal components. The projected paths of tholeiitic and calc-alkaline liquids shown in Fig. 1a have been tested experimentally and confirmed with analyses of naturally occurring samples from volcanic suites related by these fractionation (with or without assimilation) trends^{28,30}. Because the melt rock from the Chicxulub structure has previously been identified as an andesite or as being andesitic, we have plotted a broad range of tholeiitic and calc-alkaline andesitic compositions^{28,31} in Fig. 1a for comparative purposes, in addition to those of anhydrite-bearing pumices^{32,33}.

The composition of the groundmass in the Chicxulub melt rock was determined by modal reconstruction of electron microprobe data (Table 1), which allowed us to avoid the contaminating effects of unmelted xenoliths and secondary mineralization inherent in the bulk X-ray fluorescence analysis⁴. In contrast to the compositions in the diverse set of volcanics, the composition of the Chicxulub groundmass plots well off the calc-alkaline and tholeiitic trends in the augite- and plagioclase-saturation field in Fig. 1b. The composition of the groundmass varies slightly from one area to another in thin section, forming a coherent mixing trend parallel to the clinopyroxene-quartz join, but each of these compositions and their average are clearly not

the products of normal chemical differentiation, and are thus the likely products of shock-induced melting during an impact event. Although compositions are not often calculated by modal reconstruction, the errors associated with the method (Table 1) are not nearly large enough to be responsible for the disparity between the projected groundmass composition and that of other igneous products. Neither is mixing of volcanic magmas likely to be responsible for the groundmass composition, because it would be difficult to produce one of the necessary endmember magmas, which must lie off the tholeiitic and calc-alkaline differentiation trends, across the augite-olivine-plagioclase cotectic, and in the augite-saturated field.

Disparities between melt compositions produced by volcanic processes and those of impact melts occur because impact melt rocks are produced by the shock-melting of the target material, which in most circumstances is a mixture of lithologies, often including sediments, that are not related by any single igneous fractionation trend. This would have certainly been the case on the Yucatán peninsula where the impact would have melted carbonates, evaporites and sandstones. Because the composition of the melt rock is inconsistent with that produced by a mixture of these types of sediments, a portion of the underlying crystalline basement must also have been melted and incorporated. This is consistent with the expected volume of melt in a crater the size of Chicxulub, which has been estimated to be at least 0.3 times the displaced volume, on the basis of a transient crater diameter of 110 km (ref. 4). Using scaling relations derived from other impact craters²⁷, one finds that this corresponds to at least 3.7×10^4 to 1.0×10^5 km³ of melt for a minimum impact velocity of 11 km s⁻¹. Larger estimates of the transient crater diameter of 130 to 144 km, calculated from the thickness of K/T boundary ejecta in the Caribbean and North America regions²⁷, suggest a minimum of 4.9×10^4 to 3.1×10^5 km³ of melt. If the impact velocity was larger, say that of a short-period comet (35 km s⁻¹), then 7.3×10^4 to 4.6×10^5 km³ of melt may have been produced in the case of the larger transient crater diameters.

Of particular interest is whether the Chicxulub impact melt is related to the impact melt spherules found in K/T boundary ejecta deposits in Haiti^{6–9}. The relic glass in these spherules is generally more calcic and less siliceous than that of most other groups of tektites, implying that there was substantial calcium carbonate and/or sulphate in the target. Sigurdsson and others⁷ and Kring and Boynton⁸ therefore suggested that they were ejected from Chicxulub. Their unusually high sulphur content

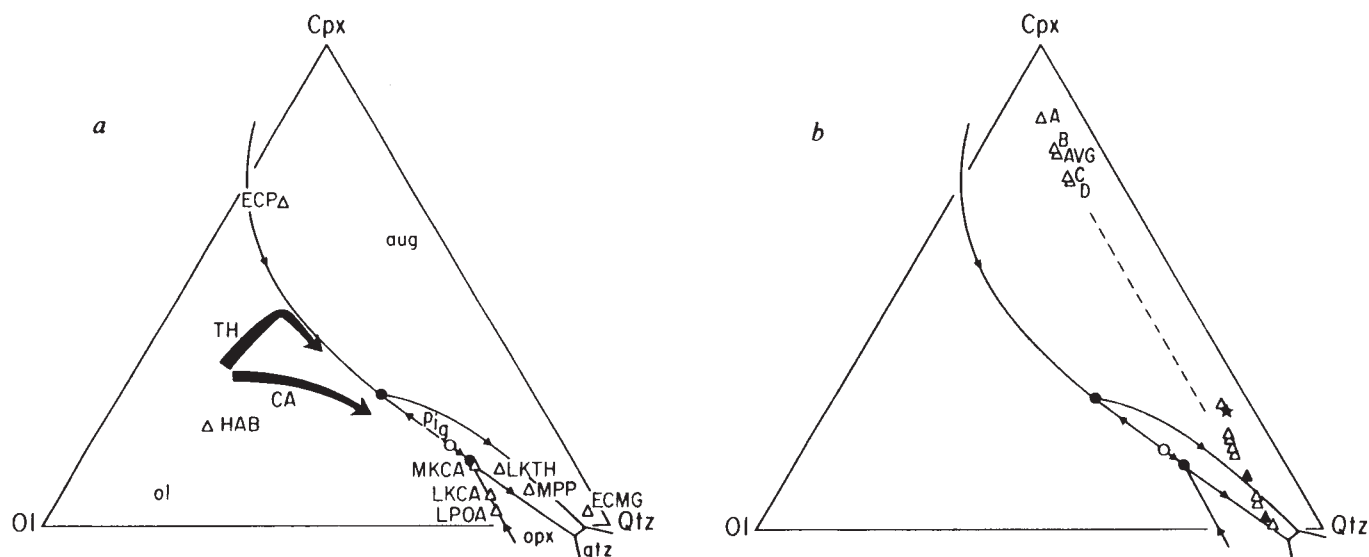


FIG. 1 *a*, Chemical trends of tholeiitic and calc-alkaline igneous suites in a pseudoternary projection of the system olivine-clinopyroxene-silica-plagioclase²⁸⁻³⁰. The arrows on the liquid line indicate the direction of decreasing temperature, the solid circles are reaction points, and the open circle is a thermal divide. Abbreviations of phases are as follows: olivine (Ol, ol), clinopyroxene (Cpx), silica (Qtz, qtz), augite (aug), pigeonite (pig), orthopyroxene (opx). Shown in this system are the compositions of a low-K tholeiitic andesite from Tonga (LKTH), a low-K calc-alkaline andesite from New Britain (LKCA), a medium-K calc-alkaline andesite from Fiji (MKCA), a platy olivine andesite from Medicine Lake (LPOA), a parental high-alumina basalt from Medicine Lake (HAB), an anhydrite-bearing pumice from Mount Pinatubo (MPP), an anhydrite-bearing pumice from El Chichon (ECP), and

matrix glass in the El Chichon pumice (ECMG)^{29,31-33}. Iron was assumed to be in the 2+ state. *b*, Composition of the groundmass of the Chicxulub melt rock sample Y6N17 (A, B, C, D and AVG) and those of the glassy K/T boundary impact melt spherules found in Beloc, Haiti (Δ Izett *et al.*⁶; ▲ Kring and Boynton⁸). The dotted line between these two populations represents a mixing trend. The composition of the Chicxulub groundmass was determined by modal reconstruction of electron microprobe analyses in this study; its average composition is listed in Table 1. Also shown is a bulk composition of the Chicxulub melt rock which was determined by X-ray fluorescence techniques (Hildebrand *et al.*⁴ (★)). All compositions are projected in molar proportions.

(G. A. Izzett), personal communication), in addition to the sulphur and oxygen isotopic composition of the glass³⁴, also seemed to suggest that they were produced from a target containing sulphates.

The compositions of the Haitian glasses are shown in Fig. 1. The projected positions of many of these glasses are discordant with the calc-alkaline and tholeiitic fractionation trends, and confirm previous evidence^{7,8} that the melt spherules are unlikely to be volcanic products, but like the groundmass of Y6N17, were probably produced by melting in an impact event. As previously noted⁷, the glasses from Haiti lie along a mixing trend. As shown in Fig. 1*b*, they also lie along the same mixing trend that was derived from composition of the groundmass in the Chicxulub melt rock, indicating that both the Haitian impact melt spherules and the Chicxulub melt rock were probably derived from the same impact event.

Because the basement lithologies of the northern Yucatán peninsula are poorly known, the exact lithologies responsible for the mixing trend between the Chicxulub melt rock and the Haitian impact melt spherules cannot be determined unambiguously. It is likely, however, that the impact melt spherules, which are more siliceous than the groundmass of the melt rock, were influenced by sandstones, shales, schists and/or granites, all of which are associated with the Yucatán peninsula^{2,35} and plot near the silica apex of Fig. 1*b*. As described above, their calcium content also indicates a contribution from calcium carbonates and/or sulphates. To demonstrate this mixing trend, we show in Table 1 that the composition of a mixture, albeit one that is non-unique, consisting of the Y6N17 groundmass, mica schist, sandstone, dolomite, and limestone or anhydrite, is very similar to the compositions of Haitian impact melt spherules.

It is less clear what lithologies were the sources of the plagioclase and clinopyroxene components of the Chicxulub melt rock. Contributions from diabbases, pyroxenites or amphibolites are possible, but the contribution from the latter could not have

been large because the resulting melt composition would then plot much closer to the olivine apex of the pseudoternary. Of particular interest is the bulk composition of the Chicxulub melt rock which was determined by X-ray fluorescence techniques and previously noted to be similar to that of the K/T spherules⁴. If one corrects for anhydrite on the basis of the measured concentration of SO₄, then its composition (Table 1) projects onto the mixing array produced by the compositions of Haitian glasses (Fig. 1*b*). Because the difference between the compositions of the groundmass reported here and that of the bulk rock determined by X-ray fluorescence is the presence of unmelted clasts, principally quartz and feldspar, it confirms the importance that their parental lithologies had on the mixing trend between the Haitian melt spherules and the groundmass of the Chicxulub melt rock. The unique mixing trend between the Haitian impact melt spherules and the Chicxulub melt rock also indicates that they were probably produced by the same impact event, and thus implies that the Chicxulub melt rock was produced precisely at the end of the Cretaceous when the impact melt spherules in Haiti were deposited. □

Received 21 October 1991; accepted 26 May 1992.

1. Cornejo Toledo, A. & Hernandez Osuna, A. *Asoc. Mex. Geol. Petrol. Bol.* **2**, 453-460 (1950).
2. Lopez-Ramos, E. in *The Ocean Basins and Margins*, Vol. 3, *The Gulf of Mexico and the Caribbean* (eds Nairn, A. E. M. & Stehli, F. G.) 257-282 (Plenum, New York, 1975).
3. Penfield, G. T. & Camargo Z., A. *Soc. Explor. Geophys. 51st Ann. Meeting Tech. Prog. (Abstr.)* **37** (1981).
4. Hildebrand, A. R. *et al. Geology* **19**, 867-871 (1991).
5. Hildebrand, A. R. & Boynton, W. V. *Science* **248**, 843-847 (1990).
6. Izzett, G. A., Maurrasse, F. J.-M. R., Lichte, F. E., Meeker, G. P. & Bates, R. *U.S. Geol. Surv. Open-file Rep.* 90-635, 31 (1990).
7. Sigurdsson, H. *et al. Nature* **349**, 482-487 (1991).
8. Kring, D. A. & Boynton, W. V. *Geochim. cosmochim. Acta* **55**, 1737-1742 (1991).
9. Maurrasse, F. J.-M. R. & Sen, G. *Science* **252**, 1690-1693 (1991).
10. Simonds, C. H., Floran, R. J., McGee, P. E., Phinney, W. C. & Warner, J. L. *J. geophys. Res.* **83**, 2773-2788 (1978).
11. Taylor, F. C. & Dence, M. R. *Can. J. Earth Sci.* **6**, 39-45 (1969).
12. Grieve, R. A. F. *Geol. Soc. Am. Bull.* **86**, 1617-1629 (1975).
13. French, B. M., Hartung, J. B., Short, N. M. & Dietz, R. S. *J. geophys. Res.* **75**, 4396-4406 (1970).
14. Fudali, R. F. *J. geophys. Res.* **79**, 2115-2121 (1974).

15. Carstens, H. *Contrib. Mineral. Petrol.* **50**, 145-155 (1975).
16. Dence, M. R. *J. geophys. Res.* **76**, 5552-5565 (1971).
17. Phinney, W. C., Simonds, C. H., Cochran, A. & McGee, P. E. *Proc. 9th lunar planet. Sci. Conf.*, 2659-2693 (1978).
18. Kieffer, S. W. & Simonds, C. H. *Rev. Geophys. Space Phys.* **18**, 143-181 (1980).
19. Reimold, W. U., Barr, J. M., Grieve, R. A. F. & Durrheim, R. J. *Geochim. cosmochim. Acta* **54**, 2093-2111 (1990).
20. Smith, A. L. & Carmichael, I. S. E. *Contrib. Mineral. Petrol.* **19**, 212-238 (1968).
21. Sato, H. *Contrib. Mineral. Petrol.* **50**, 49-64 (1975).
22. Vance, J. A. *J. Geol.* **73**, 636-651 (1965).
23. Tsuchiyama, A. *Contrib. Mineral. Petrol.* **89**, 1-16 (1985).
24. Arculus, R. J., Johnson, R. W., Chappell, B. W., McKee, C. O. & Sakai, H. J. *Volc. geotherm. Res.* **18**, 215-247 (1983).
25. Grieve, R. A. F. & Floran, R. J. *J. geophys. Res.* **83**, 2761-2771 (1978).
26. Reimold, W. U. *Geochim. cosmochim. Acta* **46**, 1203-1225 (1982).
27. Melosh, H. J. *Impact Cratering: A Geologic Process*, 245 (Oxford Univ. Press, New York, 1989).
28. Grove, T. L., Gerlach, D. C. & Sando, T. W. *Contrib. Mineral. Petrol.* **80**, 160-182 (1982).
29. Grove, T. L., Gerlach, D. C., Sando, T. W. & Baker, M. B. *Contrib. Mineral. Petrol.* **82**, 407-408 (1983).
30. Grove, T. L. & Baker, M. B. *J. geophys. Res.* **89**, 3253-3274 (1984).
31. Gill, J. *Orogenic Andesites and Plate Tectonics* (Springer, New York, 1981).
32. Lühr, J. F., Carmichael, I. S. E. & Varekamp, J. C. J. *Volc. geotherm. Res.* **23**, 69-108 (1984).
33. Bernard, A., Demaffre, D., Mattioli, N. & Punongbayan, R. S. *Nature* **354**, 139-140 (1991).
34. Sigurdsson, H. *et al. Nature* **353**, 839-842 (1991).
35. Donnelly, T. W., Horne, G. S., Finch, R. C. & López-Ramos, E. in *The Caribbean Region*, vol. H, *The Geology of North America* (eds Dengo, G. & Case, J. E.) 37-76 (Geological Society of America, Boulder, 1990).
36. Flanagan, F. J. *U.S. Geol. Surv. Prof. Pap.* **840** (1976).

ACKNOWLEDGEMENTS. We thank A. Camargo Z. and G. Penfield for providing the sample and H. J. Melosh for reviewing the manuscript.

Shear-wave splitting and small-scale convection in the continental upper mantle

L. I. Makeyeva*, L. P. Vinnik* & S. W. Roecker†

* Institute of Physics of the Earth, Moscow, Russia

† Department of Earth and Environmental Sciences, Rensselaer Polytechnic Institute, Troy, New York 12180-3590, USA

THE deformation of the upper mantle beneath a collisional belt is key to the dynamics of the belt, but this deformation is difficult to observe. Seismic azimuthal anisotropy manifested by shear-wave splitting provides the best geophysical evidence of deformation in the upper mantle. Here we use observations of shear-wave splitting to investigate deformation in the upper mantle beneath the Tien Shan in Central Asia. We find that underneath most of the Tien Shan the fast direction of azimuthal anisotropy is roughly parallel to the axis of the belt, but it deviates by nearly 90° where the upper mantle is anomalously hot. The former relationship, previously observed in other collisional belts, seems to be usual for this environment. The anomalous directions, however, provide evidence of a rising plume beneath the range, suggesting more generally that the along-strike flow of the mantle beneath collisional belts is a reflection of thermally driven convection.

Azimuthal anisotropy is a useful indicator of deformation in the mantle because the principal constituent of the mantle, olivine, is highly anisotropic and its lattice-preferred orientation is sensitive to the direction of flow¹⁻⁶. Many recent studies of azimuthal anisotropy are based on observations of shear-wave splitting in seismic phases SKS and SKKS⁷⁻¹⁷, and a similar approach is adopted here. SKS and SKKS propagate through the Earth's core and arrive at the receiver in a near-vertical direction. Their polarization is controlled by azimuthal anisotropy underneath the receiver.

The Tien Shan (Fig. 1) extends roughly east-west over a distance of ~2,500 km with the highest elevations exceeding 7,000 m. The region accreted onto Eurasia in the Palaeozoic and was stable throughout the Mesozoic¹⁸⁻²⁰. Tectonic activity resumed in the Oligocene, presumably as a consequence of the collision of India with Eurasia, and has continued to the present day. The central and western parts of the Tien Shan are separated by a prominent active right-lateral strike-slip fault, the Talasso-Fergana fault.

Geophysical studies during the past decade (see ref. 21) indicate that the parts of the Tien Shan separated by the Talasso-Fergana fault differ markedly in various properties (seismic velocities, attenuation and density) of the upper mantle. On the basis of these data, part of the recent uplift in the central Tien Shan east of the fault was attributed²¹ to a topographic swell associated with an elevated thermal anomaly in the upper mantle. The postulated difference in the styles of deformation between the two parts of the Tien Shan helps to explain a variety of near-surface features such as the differential motion along the fault²¹ and patterns of seismicity²². This interpretation is corroborated by more-recent seismic studies of the structure of the crust and the upper mantle of the region²³⁻²⁵.

The seismograms used here mainly are those recorded by a network of analog seismographs (Fig. 1) with three-component, intermediate-period Kirnos seismometers. All three components are recorded on a single sheet of paper at a rate of 60 mm min⁻¹. Selected seismograms (Table 1) were digitized repeatedly to minimize error and rotated into the (*R*, *T*) coordinate frame. Rotated records were band-pass filtered between periods of 8 and 15 s to reduce noise. The effect of anisotropy in these data is in the ellipticity of the particle motion, with the major and minor axes corresponding to the *R* and *T* directions, respectively (Fig. 2).

The records were inverted for polarization direction of the fast split wave (α) and time interval between the split waves (δt) by using the algorithm of ref. 10. The optimum pair of parameters was found by minimizing the r.m.s. difference between the observed *T* components of SKS and the theoretical *T* components corresponding to the observed *R* components. This pair is determined by a grid search over all possible values of the parameters (Fig. 3). The results for individual records may deviate from those for the whole set by 10-20° for α and 0.3-0.5 s for δt (Fig. 4). The errors in the final estimates are generally about 10° and 0.3 s, respectively.

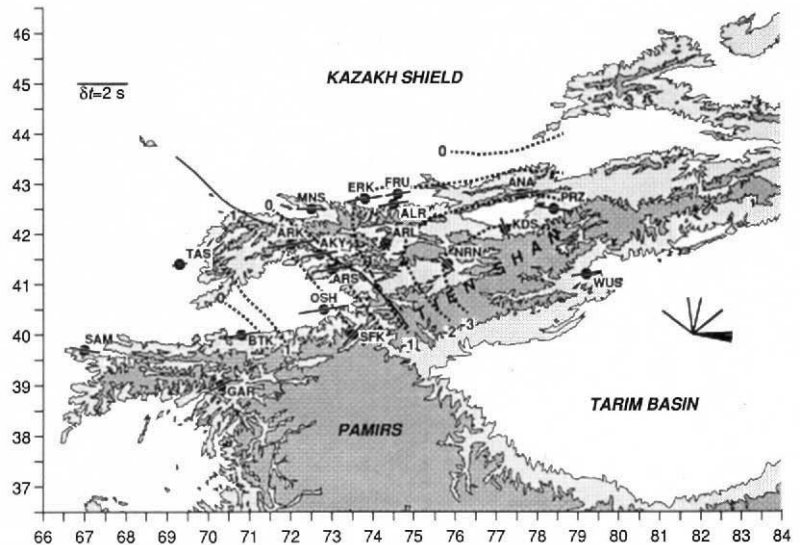
The variations of the estimates of the anisotropy parameters (Fig. 1) are fairly regular, implying that they are caused mainly by real changes in the properties of the Earth medium rather than by measurement errors. At most of the stations the fast direction is near 80°, and the largest values of δt (near 2 s) are observed in a region just to the west of the fault (stations ARK, ARS and OSH). The exceptions to this pattern are found east of the fault where the fast direction is either indeterminate (ARL) or nearly perpendicular to the normal (KDS and NRN).

TABLE 1 Directions and magnitudes of anisotropy determined here

Station	SKS back azimuths (°)	Azimuth (°)	δt (s)
TAS	48, 48, 96, 98, 99	80	0.6
SAM	12, 48, 48, 88, 99	100	1.2
GAR	12, 48, 88, 96, 98, 99, 99, 310	60	1.0
BTK	48, 88, 90, 96, 99, 99	80	1.4
ARK	12, 48, 90, 96, 96, 99, 99	80	1.8
AKY	48, 96, 99, 310, 353	80	1.4
ARS	48, 48, 88, 96, 96, 99, 99, 353	80	1.6
OSH	12, 48, 88, 90, 96, 99, 99	80	2.0
SFK	12, 48, 48, 88, 96, 96, 99	40	1.2
MNS	48, 88, 90, 96, 98, 99, 353	90	1.2
ERK	12, 48, 96, 99	80	1.4
FRU	12, 48, 90, 96, 98, 99, 310	80	1.0
ALR	48, 90, 98, 98, 99	80	1.2
ARL	88, 90, 96, 96, 99	(60)	(0.6)
NRN	12, 90, 353	140	0.8
KDS	12, 48, 90, 99, 99	160	1.0
ANA	12, 48, 48, 90, 99, 99	80	1.4
PRZ	48, 90, 98, 99, 310	100	1.6

'SKS back azimuths' refer to the station-event azimuths for SKS phases used in determining anisotropy at each station. The values in parentheses at station ARL indicate that they are poorly constrained.

FIG. 1 Directions and magnitudes of azimuthal anisotropy determined here, plotted on a map of regional topography. Locations of stations are indicated by solid circles. Lines at each station point in the direction of polarization of the fast quasi-shear wave, and their lengths are proportional to the delay between the quasi-shear waves (δt). The cross at ARL indicates the indeterminacy of anisotropy at that station. Lines at right-hand side indicate back azimuths to the 14 events used here. Thick dashed lines are contours of the P-wave velocity deviations from an average velocity in the depth range of 50–100 km determined by inversion of arrival times from teleseismic events²³. Dark shading corresponds to elevations greater than 3,000 m, light shading to elevations between 1,500 and 3,000 m, and no shading to elevations less than 1,500 m. The thick line passing just to the east of stations AKY, ARK, and OSH shows the position of the Talasso-Fergana fault. Station WUS is a GEOSCOPE station and the anisotropy shown is that reported by ref. 16.



The amount of data for ARL is similar to those at the other stations and the lack of a reliable solution for α is caused by a small value of δt . This small δt may reflect a zone of transition between the two main patterns of anisotropy because the cumulative effect of superimposed anisotropies with orthogonal directions would be small. Note that the anomalous region coincides with the central part of the upper-mantle thermal anomaly as defined by the slowest P velocities (Fig. 1).

First we must ask whether the observed anisotropy is related to present-day processes or is inherited from the past. Significant preferred orientation of olivine can be created by deformation only at temperatures higher than $\sim 1,100$ K (refs 3, 6, 26). This threshold roughly corresponds to the thermal boundary between the lithosphere and the asthenosphere²⁶, and therefore anisotropy in the lithosphere is most probably 'frozen' from the past.

It has been suggested²⁷ that transcurrent deformation in the subcrustal lithosphere of collisional belts is the main source of anisotropy observed in these regions. A good example of this type of deformation in the Tien Shan is the Talasso-Fergana fault zone, striking northwest-southeast (Fig. 1). Shear flow in the mantle beneath and along this fault would align the fast polarization direction with the strike of the fault²⁻⁶. The absence of any evidence of this alignment in our data suggests that this effect, if present, is negligible.

The 80° direction observed at most stations is aligned with the axis of the belt and is perpendicular to the direction of convergence between India and Eurasia. A similar alignment has been found in other collisional belts (for example, the Balkans, the Caucasus, the Hindu Kush, the Himalayas and the Andes, although a solitary station in Burma, CHTO, provides a possible exception to this rule^{13-15,17}). From these observations, we infer that the 80° azimuth for the fast axis is 'normal' for the Tien Shan; that is, it reflects the active deformation processes at work in the collisional belt. Thus, either there is no significant anisotropy frozen in the lithosphere, or the fast direction of the frozen anisotropy is the same as that in the underlying asthenosphere.

A simple explanation of this alignment derives from the results of deforming ultramafic rocks by uniaxial compression²⁸, where it is shown that the [010] axis of olivine aligns with the direction of compression while the two other ([001] and [100]) axes are oriented randomly in the perpendicular plane. If compression is horizontal, then the polarization of the slow split wave of SKS is roughly parallel to the [010] axis, whereas the fast split wave is polarized roughly perpendicular to this direction. The correlation of the slow direction with the convergence direction

in collisional belts suggests that the direction of shortening in these belts is usually the same in the crust and the upper mantle.

The enhanced mobility of olivine at high temperatures (more than $\sim 1,100$ K) implies that the anisotropy in the hot upper mantle beneath the central Tien Shan is less likely to be influenced by frozen anisotropy than the relatively cold, 'normal' upper mantle to the west. As it seems that frozen anisotropy has no significant signature in the west, it is unlikely that it does in the east. We infer, therefore, that the anomalous fast directions of azimuthal anisotropy in the central Tien Shan are due to present-day processes, and that the deformation in the upper mantle here is not, as in other parts of the belt, a passive response to compression brought on by plate convergence. The simplest,

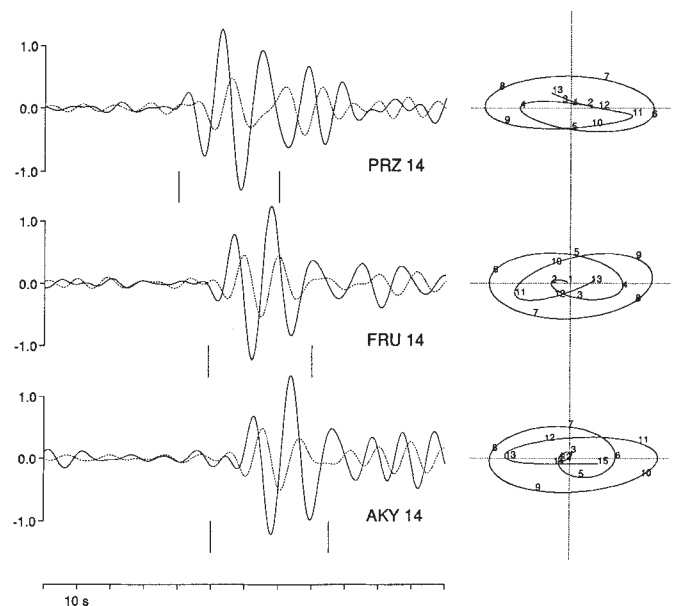


FIG. 2 Examples of SKS arrivals at three stations from the same earthquake. The seismograms shown by solid lines correspond to radial components (the horizontal components of motion radially away from the source) and the dashed lines to transverse components (the horizontal components of motion normal to the radial). Amplitudes are normalized to the radial component. Particle motions for SKS are shown on the right. The start and end points for computation of particle motion are indicated by the vertical bars beneath each pair of seismograms. In each case the radial and transverse components are on the horizontal and vertical axes, respectively. Numbers on the motion plots correspond to every 2.5 s of record.

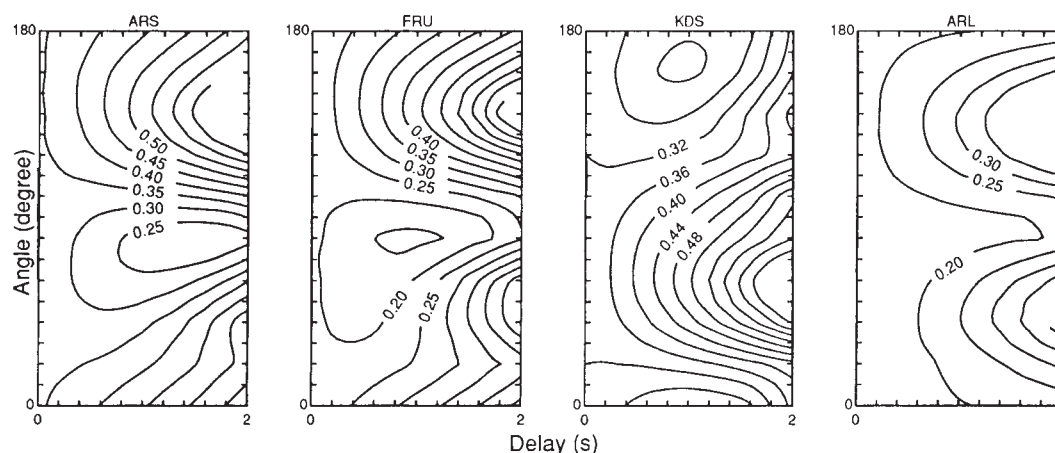


FIG. 3 Examples of contour plots of the root-mean-square misfit between calculated and observed transverse components normalized to the energy

of the radial component. The optimal values of azimuth (α) and delay time (δt) are those for which this function is minimized.

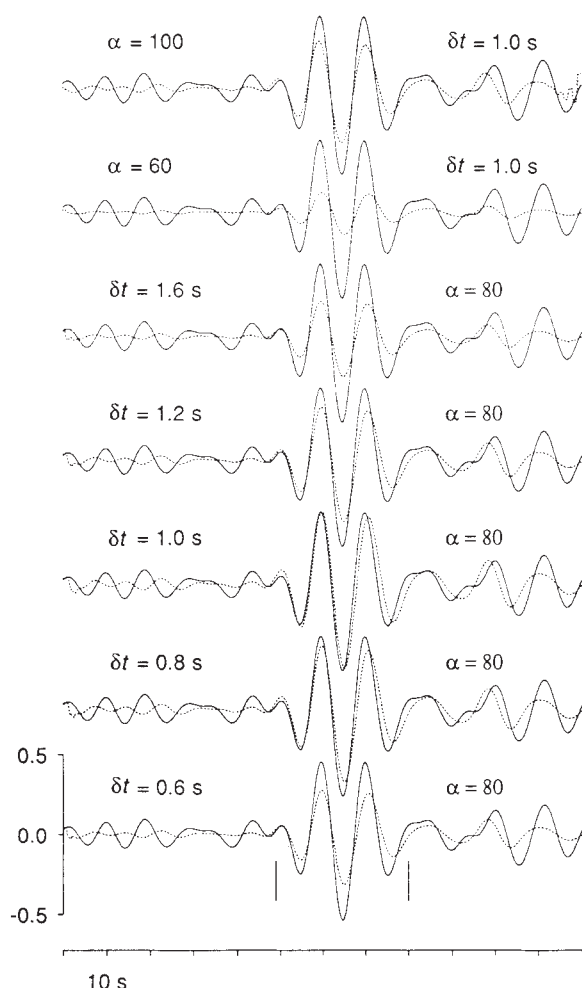


FIG. 4 Example of the sensitivity of constructed transverse components to choices of azimuth (α) and delay time (δt) at station FRU. In each case the solid line corresponds to the observed transverse component, and the dashed line to the transverse component constructed from the radial component for different choices of α and δt . Each example shows the result of deviating from the optimal values $\alpha = 80^\circ$ and $\delta t = 1.0$ s. The upper two examples vary α while keeping δt at 1.0 s, whereas the lower five vary δt while keeping α at 80° . The example with $\delta t = 1.0$ s corresponds to the optimal choice.

and perhaps only, explanation of flow contrary to the north-south convergence between India and Eurasian is small-scale thermal convection associated with the hot upper mantle. If the flow is in the form of simple shear, then the fast direction of anisotropy coincides with the direction of flow^{1-3,5}. The directions of flow thus inferred for the anomalous region suggest a spreading mantle diapir²⁹ or an inflating mantle plume³⁰ whose conduit is located somewhere in the hot mantle beneath the central Tien Shan, centred perhaps in the -3% velocity contour shown in Fig. 1.

The development of the plume could be related to instability caused by lithospheric thickening in the region of convergence, a phenomenon examined numerically in ref. 31. Sinking cold material draws the thermal boundary layer into the downwelling plume, which in turn draws hot material from below into the upwelling plume. The most likely location for the upwelling plume would be a region where the temperature before thickening was relatively high. The cold, downwelling plume has been postulated³² to lie beneath the eastern Tien Shan (outside the region in Fig. 1) whereas for the central Tien Shan there is evidence of an elevated prethickening temperature³³. These conditions thus favour convective flow between the central and eastern Tien Shan, with a distance of ~ 400 km between upwelling and downwelling plumes. In this scenario, the upwelling plume would be located beneath, and probably to the south of, the stations reporting anomalous flow directions (KDS, NRN and perhaps ARL), and flow to the east would occur some distance to the east of the network.

Finally, we note that the alignment of the fast polarization direction with the strike of the Tien Shan at most stations could be interpreted as an indication of shear flow underneath and along the range. This flow might represent near-horizontal parts of convection cells connecting regions of localized downwelling and upwelling in the mantle. By analogy, such cells could be a common feature of the mantle beneath collisional belts. $\square$

Received 20 February; accepted 27 May 1992.

- Nicolas, A. & Poirier, J. P. *Crystalline Plasticity and Solid State Flow in Metamorphic Rocks*, 444 (Wiley, London, 1976).
- McKenzie, D. *Geophys. J. R. astr. Soc.* **58**, 687-715 (1979).
- Nicolas, A. & Christensen, N. I. *Formation of Anisotropy in Upper Mantle Peridotites: a Review* (Am. Geophys. Un., Washington DC, 1987).
- Ribe, N. M. *J. geophys. Res.* **94**, 4213-4223 (1989).
- Ribe, N. M. & Yu, Y. *J. geophys. Res.* **96**, 8325-8335 (1991).
- Babuska, V. & Cara, M. *Seismic Anisotropy in the Earth* (Kluwer, Dordrecht, 1991).
- Vinnik, L. P., Kosarev, G. L. & Makeyeva, L. I. *Dokl. Akad. Nauk USSR* (in Russian) **278**, 1335-1339 (1984).
- Kind, R., Kosarev, G. L., Makeyeva, L. I. & Vinnik, L. P. *Nature* **318**, 358-361 (1985).
- Silver, P. G. & Chan, W. W. *Nature* **335**, 34-39 (1988).
- Vinnik, L. P., Farra, V. & Romanowicz, B. *Bull. seism. Soc. Am.* **79**, 1542-1558 (1989).

11. Makeyeva, L. I., Plesinger, A. & Horalek, J. *Phys. Earth planet. Inter.* **623**, 298–306 (1990).
12. Savage, M. K., Silver, P. G. & Meyer, R. P. *Geophys. Res. Lett.* **17**, 21–24 (1990).
13. Silver, P. G. & Chan, W. W. *J. geophys. Res.* **96**, 16429–16454 (1992).
14. Vinnik, L. P., Usenko, A. Yu., Makeyeva, L. I. & Oreshin, S. I. *Dokl. Akad. Nauk USSR* **318**, 1132–1136 (1991).
15. Milev, A. & Vinnik, L. P. *Dokl. Akad. Nauk USSR* **318**, 1345–1349 (1991).
16. Farra, V., Vinnik, L. P., Romanowicz, B., Kosarev, G. L. & Kind, R. *Geophys. J. Int.* **106**, 421–432 (1991).
17. Vinnik, L. P., Makeyeva, L. I., Milev, A. & Usenko, A. Yu. *Geophys. J. Int.* (submitted).
18. Burtman, V. S. *Am. J. Sci.* **A275**, 157–186 (1975).
19. Kravchenko, K. N. in *Geodynamics of Pakistan* (eds Farah, A. & de Jong, K. A.) 27–40 (Geological Survey of Pakistan, Quetta, 1979).
20. Krestnikov, V. N. *History of Development of Oscillatory Movements of the Earth's Crust of the Pamir and Neighboring Parts of Asia* (in Russian) (Academy of Sciences Publishing, Moscow, 1962).
21. Vinnik, L. P. & Saipbekova, A. M. *Ann. geophys.* **2**, 621–626 (1984).
22. Junga, S. L. *Methods and Results of Investigations of Seismotectonic Deformations* (in Russian) 191 (Nauka, Moscow, 1990).
23. Burmakov, Yu. A., Vinnik, L. P., Saipbekova, A. M. & Treusov, A. V., in *Dynamics and Evolution of the Lithosphere*, 90–100 (Moscow, Nauka, 1986).
24. Kosarev, G. L., Petersen, N. V., Vinnik, L. P. & Roecker, S. W. *J. geophys. Res.* (submitted).
25. Roecker, S. W. *et al. J. geophys. Res.* (submitted).
26. Estey, L. H. & Douglas, B. J. *J. geophys. Res.* **91**, 11393–11406 (1986).
27. Vauchez, A. & Nicolas, A. *Tectonophysics* **185**, 183–201 (1991).
28. Christensen, N. I. & Crosson, R. S. *Tectonophysics* **6**, 92–102 (1968).
29. Artush, E. V. *Geodynamics* (in Russian) (Nauka, Moscow, 1979).
30. Olson, P. in *The Encyclopedia of Solid Earth Geophysics*, 788–802 (Van Nostrand Reinhold, New York, 1989).
31. Houseman, G. A., McKenzie, D. P. & Molnar, P. *J. geophys. Res.* **86**, 6115–6132 (1981).
32. Burov, E. V., Kogan, M. G., Lyon-Caen, H. & Molnar, P. *Earth planet. Sci. Lett.* **98**, 367–383 (1990).
33. Knäuf, V. I., Micolalchik, A. V. & Christov, E. V. in *Seismotectonics and Seismicity of the Tien Shan* (ed. Kalmurzayev, K. E.) 3–18 (Frunze, Ilim, 1980).

Queen activation of lazy workers in colonies of the eusocial naked mole-rat

Hudson K. Reeve*

Section of Neurobiology and Behaviour, Seeley G. Mudd Hall, Cornell University, Ithaca, New York 14853 USA

* Present address: Harvard University, Museum of Comparative Zoology, 26 Oxford Street, Cambridge, Massachusetts 02138-2902, USA

EVOLUTIONARY conflicts of interest are expected to arise in genetically diverse social groups¹. In eusocial insect societies, a potential conflict exists between the queen and her workers over how active the workers should be^{2–5}, and evidence exists that queen aggression increases activity levels of her lazier workers^{2,3}. Here I provide evidence that queen aggression (shoving) in laboratory colonies of the eusocial mammal, the naked mole-rat (*Heterocephalus glaber*), is a convergently evolved manifestation of queen-worker conflict over worker activity. Queen-initiated shoves activate inherently lazy workers, which tend to be larger and/or less related to the queen than are infrequently shoved, industrious workers. In addition, queen removal selectively depresses the activity of workers that are larger and less related to her. Finally, queen shoving and worker inactivity are pronounced when colonies are satiated but not when colonies are hungry, indicating that the underlying 'work-conflict' is highly context-specific.

Naked mole-rats are subterranean, colonial rodents whose social system resembles that of the eusocial insects: a single breeding female (queen) and her one to three mates produce all of the young, which help maintain and defend the colony as do workers in the social insects^{6–8}. Although mean intra-colony relatedness is high because of extensive inbreeding in *H. glaber*⁹, a variety of aggressive behaviours occur within naked mole-rat colonies¹⁰. The most striking of these is 'shoving', in which one mole-rat violently pushes another, often hissing at and (occasionally) biting the recipient⁵. The queen is the principal initiator of shoves, directing them primarily toward larger or less related colony-mates, without regard to sex⁵. The lack of sex bias in the queen's shoving, and the fact that she frequently shoves her mates, indicate that queen-shoving is not merely a mechanism for suppressing reproductive challenges by subordinate females.

TABLE 1 Effect of colony hunger level on mean colony work level and rate of shoving by the queen ($N=6$ colonies)

	Time since last feeding (h)		
	5–15	15–25	25–45
Mean colony work level	0.06 (0.02)*	0.07 (0.02)†	0.13 (0.03)*†
Queen's shove rate (shoves/h)	5.4 (1.0)*	3.7 (0.3)	2.4 (1.3)*

Mean work level of colony members (excluding the queen) and the queen's shove rate, each as a function of hunger level (undisturbed colonies). All observations occurred between 10:00 and 16:00 h (1985–1990). I observed three colonies at Cornell University (A, K and TT) and three at the University of Michigan (A(M), B(M) and C(M)). Descriptions of the tunnel layouts, demographics, histories, maintenance procedures, diets, compositions, and observation schedules for these colonies are provided elsewhere^{5,14}. Mean queen-worker relatedness (as determined by DNA fingerprint data⁹ together with a knowledge of colony genealogies^{5,15}), mean worker body weight, and colony size (minus queen) for the six colonies, respectively, are: colony A (0.75, 37.2 g, 27); colony K (0.59, 30.7 g, 20 in 1985–6 and 0.19, 37.1 g, 17 in 1988 after an unrelated female replaced the old queen); colony TT (0.81, 32.3 g, 31); colony A(M) (0.77, 27.7 g, 25); colony B(M) (0.63, 29.5 g, 27); colony C(M) (0.47, 34.7, 19). Variability in kinship resulted from variable numbers of the queen's uncles, aunts, siblings and offspring as well as presumed unrelated individuals within colonies, simulating turnover of breeders within and occasional immigration of unrelated conspecifics into field colonies (indicated by genetic data⁹). Behavioural states of all individuals in the tunnel system were sampled by taking instantaneous scans of the tunnel system at 5-min intervals for 0.5–1.0 h per session. (5,320 individual scans; 110 h observations at 5–50 h food-deprivation intervals.) Work level is the proportion of scans in which a worker was in a work state.

* or †, $P < 0.05$; two-tailed, Friedman test. There was no sex bias in the shoving at any level of food deprivation; $P > 0.10$.

According to the work-conflict hypothesis, a high-ranking worker is a 'hopeful reproductive'¹¹ and should be lazy, because laziness reduces its energy expenditure or exposure to predation, hence increases the probability that it, rather than one of its fellow workers, will replace a breeder. The latter outcome is favourable for the lazy worker (whether male or female), because it is always more closely related to its own offspring than to that of another worker. The queen, by contrast, benefits maximally if all her workers are active, because she is always more related to her own offspring than to any worker's offspring, and worker activity benefits her own offspring. The predicted result is queen-worker conflict over worker activity levels.

In naked mole-rats, relatively large or rapidly growing individuals appear more likely to acquire and maintain breeding status^{5,12}, and individuals less related to the queen (called 'less related workers') have less to gain by working for her, hence both classes of mole-rats are expected to be relatively lazy according to the work-conflict hypothesis, especially in light of evidence that tasks such as tunnelling exposes mole-rats to predation by snakes in the field^{6,13}. Male breeders will be in work-conflict with the queen if multiple paternity reduces each breeding male's relatedness to offspring compared to the queen's relatedness to offspring.

I tested two critical predictions of the work-conflict hypothesis as applied to queen-shoving in naked mole-rats. The hypothesis predicts that (1) shoving should increase the work rates of the recipients of shoves, and (2) the most frequent targets of queen-shoving should work at the lowest rates, at least in the absence of the queen's influence.

The colony's workers exhibit a variety of behaviours in their tunnel system, including digging, chewing on substrate, sweeping debris, carrying food or nest material, walking, urinating, eating tuber, grooming and being immobile (including sleeping in the nestbox)¹⁰. Only the first four behaviours are defined as 'work' states. The work states appear to be related to exploration for food, as occurs in the field⁶. As colonies became more food

TABLE 2 Evidence that queen shoving activates workers ($N=6$ colonies)

(a)	Before shove	After shove
Shove-recipient's work level		
All shoves	0.14 (0.03) [†]	0.25 (0.06) [†]
Tunnel shoves only	0.34 (0.05) [†]	0.58 (0.07) [†]
(b)	Queen removed	Control
Work level (satiated colonies)	0.049 (0.02)*	0.077 (0.02)*
Work level (hungry colonies)	0.110 (0.02)	0.084 (0.02)
(c)	More work after queen removal	Less work after queen removal
Queen's rate of shoving toward recipient (shoves h^{-1})	0.033 (0.02)*	0.167 (0.03)*

a, Overall mean work levels of workers that were shoved by the queen in scans immediately before and after the shoves (undisturbed colonies; standard errors in parentheses). Between scans, I observed the queen, who periodically patrolled the tunnel system (once per 16 min) and often shoved workers in tunnels⁵. Statistical tests are two-tailed, Wilcoxon matched pairs, signed-ranks tests (as in b and c below). Pairs of values indicated by * or † are significantly different at the 0.05 level and 0.01 level, respectively (as below). The mean work level in the second scan after the shove (0.22) was slightly less than that in the first post-shove scan (0.25), but still significantly greater than that for the scan preceding the shove (0.14; $P<0.01$). Thus, the effect of the queen's shove lasts at least 1/3 of the time between successive queen shoves⁵ or tunnel patrols. b, Mean work levels of all workers 4 h after removal of the queen ('queen removed' treatments) and 4 h after removal, then immediate replacement, of the queen ('controls'). Observations were conducted for 1 h. The 4-h latency allowed colony activity to equilibrate after the initial disturbance. (Treatments: 660 individual scans; 15 observation-hours; controls: 770 scans; 16 observation-hours.) Results are presented both for very satiated colonies (0–5 h food deprivation) and for very hungry colonies (40–50 h food deprivation). Control and treatment observations occurred in random order, separated by at least 3 days. Replicate 1986 data from colony K were averaged as means. c, Comparison of the rates at which shoves were received from the queen for individuals that increased ('more work') or decreased ('less work') their work levels after queen removal (satiated colonies).

deprived, the mean work level of mole-rats significantly increased (Table 1). Moreover, the mean correlation between individual weight and work level, averaged across all six colonies, is significantly negative ($\bar{r} = -0.38$, $P<0.02$, two-tailed, one-sample t -test (null $\bar{r}=0$)), indicating that smaller workers forage at higher rates, as also occurs in the field⁶.

The queen's total rate of shoving significantly declined with increased food deprivation interval (Table 1). Thus, as predicted by the work-conflict hypothesis, the queen shoves at a higher rate when there are fewer active workers. A comparison of mean work levels for the same workers immediately before and after receiving shoves from the queen reveals that shoving produced a nearly twofold increase in the rate of working (Table 2a). Shoves initiated in the nestbox were directed either at returning mole-rats lacking foraged items (a mean of 33% of such returns elicited shoves by the queen) or at inactive mole-rats, who, in a mean of 17% of these cases, then immediately entered tunnels to work. Shoves in tunnels increased the work-rates of recipients (Table 2a). Transport of food by workers from the tunnels to the nestbox protected these workers from queen shoving: Out of a total of 66 such observed transports of food (tuber) fragments, in no case was the deliverer of the food shoved by the queen.

In satiated colonies (food-deprived 0–5 h), queen removal caused a significant decrease in mean work levels (Table 2b). But in hungry colonies (food-deprived 40–50 h), queen removal had no significant effect on mean work level (Table 2b). Thus, queen removal caused significant reductions of work level, but only in a context (satiation) that is associated in queenright

TABLE 3 Evidence that larger workers and workers less related to the queen exhibited greater drops in work level when the queen is removed from satiated colonies

	Mean weight (g) of active workers		Mean relatedness to queen of active workers	
	Queen removed	Control	Queen removed	Control
All colonies	28.8 (2.1)*	31.4 (2.6)*	0.81 (0)	0.75 (0.1)
Colony K only	26.0 (1.7) [†]	32.1 (0.5) [†]	0.78 (0.03) [†]	0.42 (0.07) [†]

Mean relatedness of active workers to the queen and mean weight of active workers both when the queen was present (controls) and 4 h after the queen was removed (* $P<0.05$, Wilcoxon matched-pairs, signed-ranks test across all six colonies). Mean body weight of active workers is equal to $\sum a_i W_i / \sum a_i$, where a_i is the work level and W_i is the weight of the i th worker. Mean relatedness of active workers is equal to $\sum a_i G_i / \sum a_i$, where a_i is the work level and G_i is the relatedness to the queen for the i th worker. The selective decline in the work levels of larger, less related workers did not occur in hungry colonies ($P>0.10$ for both comparisons). Colony K only: mean relatedness of active workers to the queen and mean weight of active workers both when the queen was present (controls) and 4 h after the queen was removed for colony K, the colony with the largest variability in kinship. There were three queen removal replicates and four control replicates (1986 data); treatments were separated by at least 3 days. Statistical tests are one-tailed Mann-Whitney tests ($^{\dagger}P<0.02$). Standard errors are in parentheses.

colonies with low worker activity and high rates of queen-shoving.

Larger or less related individuals tend to be lazier. As noted above, smaller animals work harder. Across colonies, mean queen-worker relatedness is significantly positively correlated with mean work levels, but only when colonies are satiated (Fig. 1). Within colonies there is a significant, positive correlation between a queen's relatedness to a worker and that worker's work level [$\bar{r} = +0.38$, $P<0.01$, two-tailed, one-sample t -test; $N=6$ colonies]. The mean within-colony correlation between the queen's shove rate and the work levels of recipients is slightly, but significantly, negative ($\bar{r} = -0.15$, $P<0.05$).

The weakness of the mean intra-colony association between work level and rate of shoves from the queen might result from the queen's preferential shoving of inherently lazier workers,

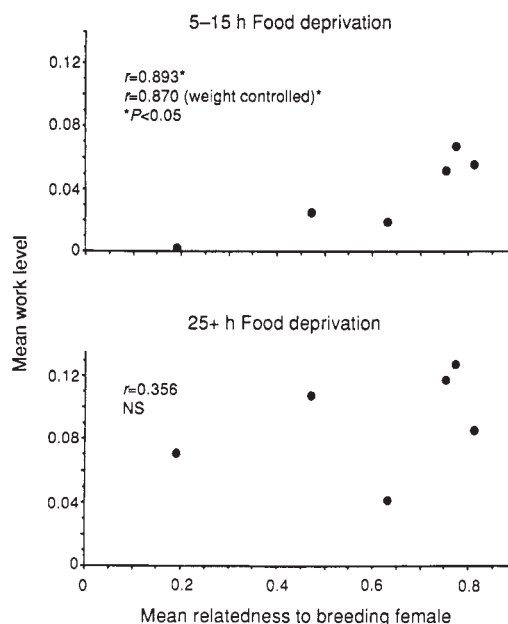


FIG. 1 Pearson correlation between overall mean colony work level and mean queen-worker relatedness for observations of the six undisturbed colonies (1988 observations for colony K were used to maximize range of relatednesses). a, Correlation for satiated colonies (5–15 h food deprivation). Partial correlation between work level and relatedness, controlling for body weight, is also shown. b, Correlation for hungry colonies (25–50 h food deprivation). Observations were made without knowledge of queen-worker relatedness. NS, not significant at 0.05 level.

that is, lazy workers might be stimulated by the queen to become nearly as active as industrious workers. Thus, worker laziness should be revealed when the queen's influence is removed. Indeed, a link between queen shove-rate and intrinsic work levels of shove-recipients is strongly indicated by a comparison of individuals that decreased versus individuals that increased their work levels after queen removal (satiated colonies): In the queen's presence, the former were shoved at nearly five times the rate of the latter (Table 2c). Moreover, when the queen was removed from satiated colonies, the mean relatedness to the queen of active workers increased, whereas the mean weight of active workers decreased (Table 3).

These data strongly support the work-conflict hypothesis for the significance of queen shoving behaviour in naked mole-rats. Thus, the function of queen aggression in naked mole-rats is similar to that in social insects with relatively small colonies, such as paper wasps^{2,3}. The naked mole-rat queen is in conflict with her larger, less related colony-mates over how much work these individuals should do, and she activates these individuals through her shoving. Breeding males might also be expected to activate subordinates; indeed, these males initiate nearly all of the few shoves not initiated by the queen⁵. These conflicts occur despite a high mean intra-colony relatedness⁹, demonstrating that even very close kinship does not preclude the appearance of conflict in animal societies.

The queen-worker conflict is highly context-specific. It seems likely that when the colony is hungry, it is in everyone's selfish interest to search for food, and thus the work conflict disappears. When colonies are relatively satiated, however, there seems to be disagreement between the queen and her larger, less related workers over how active those workers should be. At times of satiation, workers can, with lowered risk of starvation, pursue

selfish reproductive options by being lazy. But the queen would benefit more if these lazy workers continued participation in colony work (for example, searching for scattered, new food items⁶). The resulting context-dependence of the degree of queen-worker conflict demonstrates that it is unwise to categorize entire animal societies as conflictual or cooperative without specifying particular social contexts. □

Received 6 February; accepted 15 May 1992.

1. Hamilton, W. D. *J. theor. Biol.* **7**, 1-52 (1964).
2. Reeve, H. K. & Gamboa, G. *J. Behav. Ecol. Sociobiol.* **13**, 63-74 (1983).
3. Reeve, H. K. & Gamboa, G. *J. Behaviour* **102**, 147-167 (1987).
4. Schmid-Hempel, P. *Am. Nat.* **135**, 501-526 (1990).
5. Reeve, H. K. & Sherman, P. W. in *The Biology of the Naked Mole-rat* (eds Sherman, P. W., Jarvis, J. U. M. & Alexander, R. D.) 337-357 (Princeton Univ. Press, Princeton, 1991).
6. Brett, R. A. in *The Biology of the Naked Mole-rat* (eds Sherman, P. W., Jarvis, J. U. M. & Alexander, R. D.) 97-184 (Princeton Univ. Press, Princeton, 1991).
7. Jarvis, J. U. M. *Science* **212**, 571-573 (1981).
8. Jarvis, J. U. M. & Bennett, N. C. in *The Biology of the Naked Mole-rat* (eds Sherman, P. W., Jarvis, J. U. M. & Alexander, R. D.) 66-96 (Princeton Univ. Press, Princeton, 1991).
9. Reeve, H. K., Westneat, D. W., Noon, B. A., Sherman, P. W. & Aquadro, C. F. *Proc. natn. Acad. Sci. U.S.A.* **87**, 2496-2500 (1990).
10. Lacey, E. A., Alexander, R. D., Braude, S. H., Sherman, P. W. & Jarvis, J. U. M. in *The Biology of the Naked Mole-rat* (eds Sherman, P. W., Jarvis, J. U. M. & Alexander, R. D.) 209-242 (Princeton Univ. Press, Princeton, 1991).
11. West-Eberhard, M. J. in *Natural Selection and Social Behaviour* (eds Alexander, R. D. & Tinkle, D. W.) 3-17 (Chiron, New York, 1981).
12. Jarvis, J. U. M. in *The Biology of the Naked Mole-rat* (eds Sherman, P. W., Jarvis, J. U. M. & Alexander, R. D.) 384-425 (Princeton Univ. Press, Princeton, 1991).
13. Braude, S. H. in *The Biology of the Naked Mole-rat* (eds Sherman, P. W., Jarvis, J. U. M. & Alexander, R. D.) 185-194 (Princeton Univ. Press, Princeton, 1991).
14. Lacey, E. A. & Sherman, P. W. in *The Biology of the Naked Mole-rat* (eds Sherman, P. W., Jarvis, J. U. M. & Alexander, R. D.) 275-336 (Princeton Univ. Press, Princeton, 1991).
15. Reeve, H. K. thesis, Cornell Univ. (1991).

ACKNOWLEDGEMENTS. I thank C. Aquadro, A. Bass, G. Eickwort, G. Gamboa, D. Pfennig, T. Seeley, P. Sherman and J. Shellman-Reeve for comments on the manuscript, and P. Sherman and R. Alexander for allowing me to work in their laboratories. H.K.R. was supported in part by a US NSF grant to P. Sherman and C. Aquadro, and in part by a Junior Fellowship from the Harvard University Society of Fellows.

Multiple dopamine D4 receptor variants in the human population

Hubert H. M. Van Tol*†, Caren M. Wu*†, Hong-Chang Guan*†, Koichi Ohara*†, James R. Bunzow‡, Olivier Civelli‡, James Kennedy§, Philip Seeman†, Hyman B. Niznik*† & Vera Jovanovic*†

Laboratory for * Molecular Neurobiology and § Molecular Genetics, Clarke Institute of Psychiatry, 250 College Street, Toronto, Ontario M5T 1R8, Canada

† Departments of Pharmacology and Psychiatry, University of Toronto, Medical Sciences Building, Toronto, Ontario M5S 1A8, Canada

‡ Vollum Institute for Advanced Biomedical Research, Department of Cell Biology and Anatomy, Oregon Health Sciences University, Portland, Oregon 97201, USA

THE dopamine D4 receptor structurally and pharmacologically resembles the dopamine D2 and D3 receptors¹⁻⁵. Clozapine, an atypical antipsychotic that is relatively free of the adverse effects of drug-induced parkinsonism and tardive dyskinesia^{6,7}, binds to the D4 receptor with an affinity 10 times higher than to the D2 and D3 receptors¹. This may explain clozapine's atypical properties. Here we report the existence of at least three polymorphic variations in the coding sequence of the human D4 receptor. A 48-base-pair sequence in the putative third cytoplasmic loop of this receptor exists either as a direct-repeat sequence (D4.2), as a fourfold repeat (D4.4) or as a sevenfold repeat (D4.7). Two more variant alleles were detected in humans. Expression of the complementary DNA for the three cloned receptor variants showed different properties for the long form (D4.7) and the shorter forms (D4.2, D4.4) with respect to clozapine and spiperone binding. To our knowledge, this is the first report of a receptor in the

catecholamine receptor family that displays polymorphic variation in the human population. Such variation among humans may underlie individual differences in susceptibility to neuropsychiatric disease and in responsiveness to antipsychotic medication.

The genomic clone of the D4 receptor contains a region with a sevenfold direct-repeat of 48 base pairs (bp) located in the putative third cytoplasmic loop (Fig. 1). This sequence occurs twice in the isolated D4 receptor cDNA clone of the neuroepithelioma SK-N-MC (Fig. 1), as previously reported¹. Screening of cDNA libraries from human pituitary and substantia nigra, however, resulted in the isolation of variant cDNA clones of the D4 receptor. The pituitary λ gt10 clone contained a 1.4-kilobase (kb) *EcoRI* insert, coding for intron 1 and the downstream sequence of the D4 receptor. This pituitary D4 receptor clone also contained the second intron, but the last intron was spliced out. The isolated substantia nigra λ gt11 clone contained a 600-bp *EcoRI* insert, coding for the D4 receptor, starting in the 5' site of the putative third cytoplasmic loop. Both these clones contained a fourfold repeat of the 48-bp sequence in the putative third cytoplasmic loop. The absence of conventional splice junction sites^{8,9} in the sevenfold repeat sequence of the gene would require a novel splicing mechanism to account for the existence of the different cDNA clones.

Screening of genomic libraries of different human individuals resulted in the isolation of a genomic clone of the D4 receptor with a fourfold direct-repeated 48-bp sequence. It contained a non-interrupted sequence in the putative third cytoplasmic loop, identical to the cDNA isolated from the pituitary and substantia nigra. Restriction endonuclease and sequence analysis of the entire coding region, intron 2 and 3, and 1 kb of flanking regions of intron 1, did not reveal any other differences with the original cloned receptor gene. This indicated a genomic polymorphic variation for the putative third cytoplasmic loop region of the D4 receptor.

The three different D4 receptor sequences predict a polymor-

phic variation for the *Hin*CII-*Pst*I fragment of the D4 gene (Fig. 1a). Such a polymorphism was observed in humans and the different allelic fragments were sized at roughly 520 bp, 620 bp, 710 bp, 760 bp and 800 bp (Fig. 2). The 520-bp, 620-bp and 760-bp fragments correlate closely with the sizes of the *Hin*CII-*Pst*I fragments of the cloned D4 receptor variants with the two-, four- and sevenfold repeat sequence, respectively (Fig. 2). The presence of 710-bp and 800-bp fragments suggests that variants with a 6- and 8-fold repeat also exist.

To exclude the possibility that D4 receptor clones with the

twofold or fourfold repeated 48-bp sequence could be generated through splicing of the gene containing the sevenfold repeat sequence, we cloned the gene with the sevenfold repeat into the expression vector pCD-PS (ref. 10) (pCD-D4G, Fig. 1) and transfected this into COS-7 cells. After 3 days, cells were collected and poly(A)⁺ RNA was isolated for the construction of a cDNA library in λ ZAPII (Fig. 1). From this library, a 1.7-kb clone was isolated containing a full-length cDNA for the D4 receptor. All the postulated and observed conventional introns were spliced out of this cDNA¹, but the sevenfold 48-bp repeat

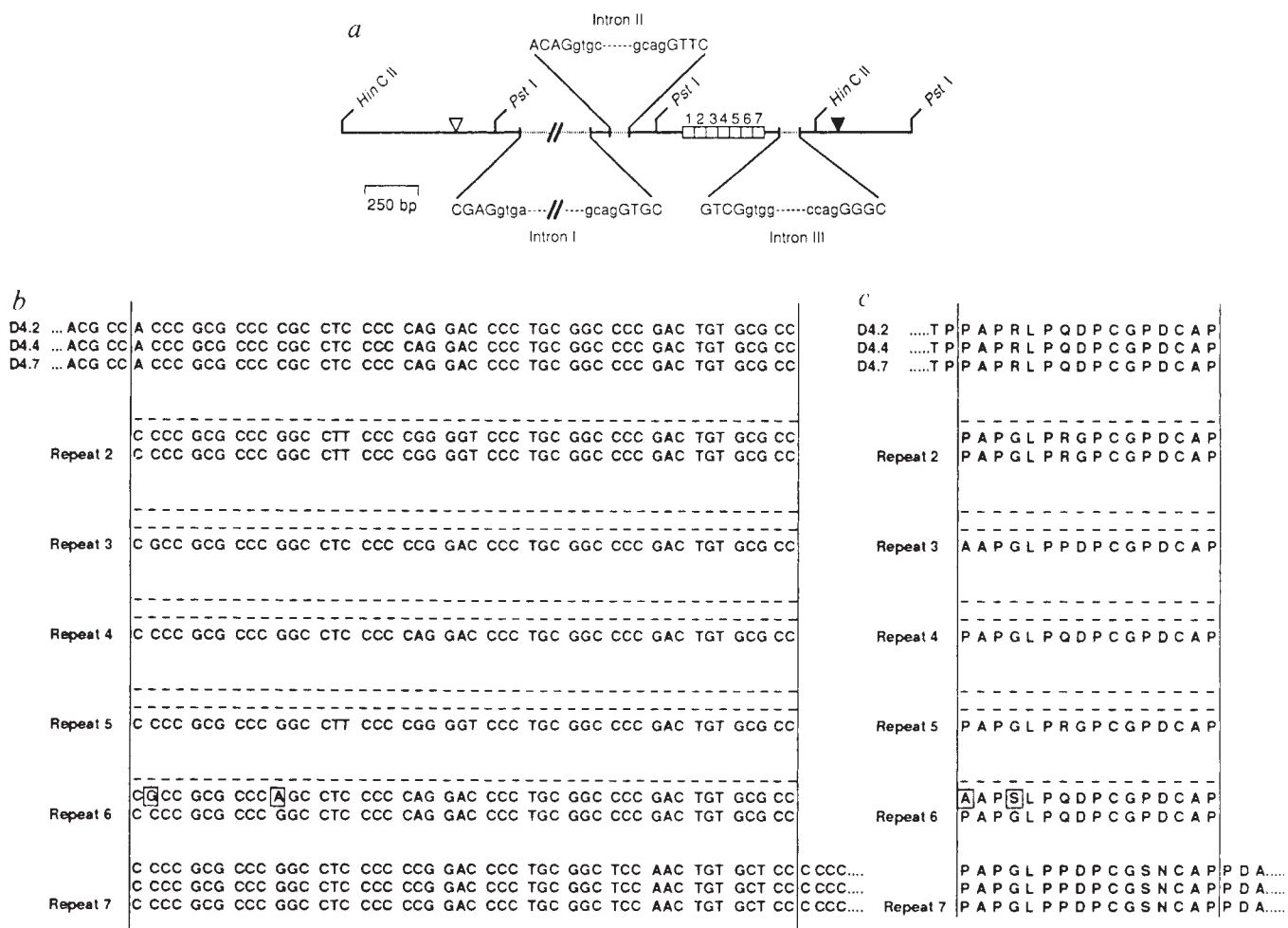


FIG. 1 *a*, Genomic organization of the human dopamine D4.4 and D4.7 receptor gene. The initiation methionine is indicated by a triangle and the stop codon by a filled triangle. The positions of the introns are indicated by broken lines. The exact location of the splice junction sites are depicted with four nucleotides of coding and intron sequence, in upper- and lower-case letters, respectively. Intron I (~2 kb) is not drawn to scale. A 48-bp sequence, which exists either as a sevenfold (D4.7) or fourfold (D4.4) or single direct-repeat sequence (D4.2), is shown as an open box numbered 1 to 7. The positions for the restriction endonuclease sites *Hin*CII and *Pst*I are indicated. *b*, Nucleotide, and *c*, amino-acid sequence for the polymorphic region of the three allelic forms of the D4 receptor. The D4.2 (ref. 1), D4.4 and D4.7 sequences are aligned for optimal homology. The dotted line indicates the gap in sequence created for optimal alignment. The homologous repeated regions are indicated on the side. Nucleotides and amino acids that vary from the D4.7, as compared to the proposed homologous regions of the other two forms, are boxed.

METHODS. Two different human genomic libraries in EMBL3 (Clontech), a human substantia nigra cDNA (Clontech), and a pituitary³ cDNA library in λ gt11 and 10, respectively, were screened for clones encoding the D4 receptor. Roughly $0.5\text{--}1 \times 10^6$ plaque-forming units (p.f.u.) were transferred in duplicate to nylon filters (DuPont/NEN) and probed with a ³²P-labelled (random primer synthesis; Boehringer Mannheim) 700-bp *Eco*RI-*Xho*I fragment encoding for the cDNA isolated from the neuroepithelioma SK-N-MC

under conditions as described before¹. Construction of pCD-D4G: a 3.8-kb *Apal*-*Pst*I fragment (containing the entire D4 gene) was isolated out of the vector pBluescript by *Kpn*I-*Xba*I digestion and subcloned into the *Kpn*I-*Spe*I sites of pCD-PS. Poly(A)⁺ RNA^{17,18} was isolated from 10⁷ COS-7 cells, which had been transiently transfected (3 days) with the expression vector containing the entire D4 gene (pCD-D4G). Roughly 5 μ g poly(A)⁺ RNA was converted into double-stranded cDNA¹⁹ (λ ZAPII cDNA synthesis kit, Stratagene), ligated into λ ZAPII arms (Stratagene), and packaged *in vitro* using Gigapack Gold packaging extracts (Stratagene). The cDNA library (300,000 p.f.u.) was plated and transferred in duplicate to nylon filters and screened for λ ZAPII cDNA clones containing D4 receptor sequences, as described previously¹. For sequence analysis, fragments from the isolated genomic clones were subcloned into the vectors pSP73 (Promega) and pBluescript (Stratagene). The isolated cDNA clones from the substantia nigra and the pituitary were excised as *Eco*RI fragments from the phage vectors and subcloned into the *Eco*RI restriction sites of pBluescript (Stratagene) and pGEM-1 (Promega), respectively, using standard techniques¹⁸. The nucleic-acid sequence of the isolated clones was determined by the dideoxy chain termination method, using 7-deaza-dGTP and Sequenase V2.0 (USB). Synthetic oligonucleotide primers (Biotechnology Service Center, Toronto) based on the D4 receptor gene sequence were used to confirm the nucleic-acid sequence of both strands of the clones.

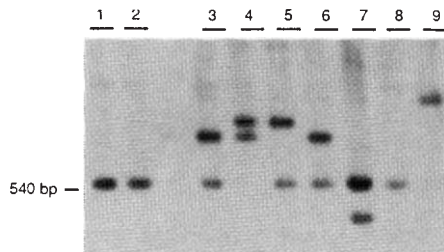


FIG. 2 Southern blot analysis of the D4 receptor gene, isolated from nine unrelated caucasian individuals (lanes 1–9) and digested with the restriction endonucleases *Pst*I and *Hin*CI.

METHODS. High-molecular-mass genomic DNA was isolated from human blood samples using proteinase K and phenol–chloroform extractions¹⁸. The genomic DNA (5 μ g) was digested with the restriction endonucleases *Hin*CI and *Pst*I and size-separated by agarose (1%) gel electrophoresis. DNA was transferred to nylon membranes (Zeta-probe, Biorad) according to standard techniques^{18,20}. The Southern blots were probed with a ³²P-labelled 600-bp *Eco*RI–*Hin*CI fragment coding for the D4 cDNA isolated from the neuroepithelioma SK-N-MC¹, and washed at high stringency (65 °C, 0.1 $\times$ SSC, 0.1% SDS, 40 min)²⁰. The blot was exposed to X-ray film for six days. The position of a 540-bp size marker is indicated on the left.

FIG. 3 Northern blot analysis of the D4 receptor transcripts of pCD-D4.7, pCD-D4.4 and pCD-D4.2. Total RNA was isolated from COS-7 cells which were transiently transfected with these constructs. Each lane contains 2 μ g total RNA. Size markers are indicated on the right. **METHODS.** Total RNA was isolated by guanidium thiocyanate/phenol/ chloroform extraction¹⁷ followed by a DNase I digestion in the presence of RNasin (Promega). RNA (2 μ g) was denatured by dimethylsulphoxide–glyoxal and size-separated by agarose (1%) gel electrophoresis²¹. The RNA was transferred and fixed to a nylon membrane (N-Bond, Amersham), and hybridized with the 700-bp *Eco*RI–*Xba*I fragment coding for the D4 cDNA from the neuroepithelioma SK-N-MC¹. For the transient expression studies, various cDNA constructs were cloned into the expression vector pCD-PS (ref. 10). A 1.7-kb *Kpn*I–*Xba*I fragment, containing the D4.7 cDNA, was cloned out of the vector pBluescript into pCD-PS and called hereafter pCD-D4.7. Full-length cDNA clones for the D4.2 and D4.4 receptor (pCD-D4.2, pCD-D4.4) were created by substituting the 920-bp *Pst*I–*Eco*RI 3' fragment of pCD-D4.7 with the 630-bp *Pst*I–*Eco*RI or 730-bp *Pst*I–*Eco*RI fragment from the D4 cDNAs isolated from the neuroepithelioma SK-N-MC and human pituitary, respectively. For transient expression COS-7 cells were transfected with the various D4 constructs by electroporation (BTX 600 Electro Cell Manipulator, 0.65 kV cm⁻¹, 4.1 ms, 5 $\times$ 10⁷ cells per ml PBS). Three days later the cells were collected and stored at –80 °C.

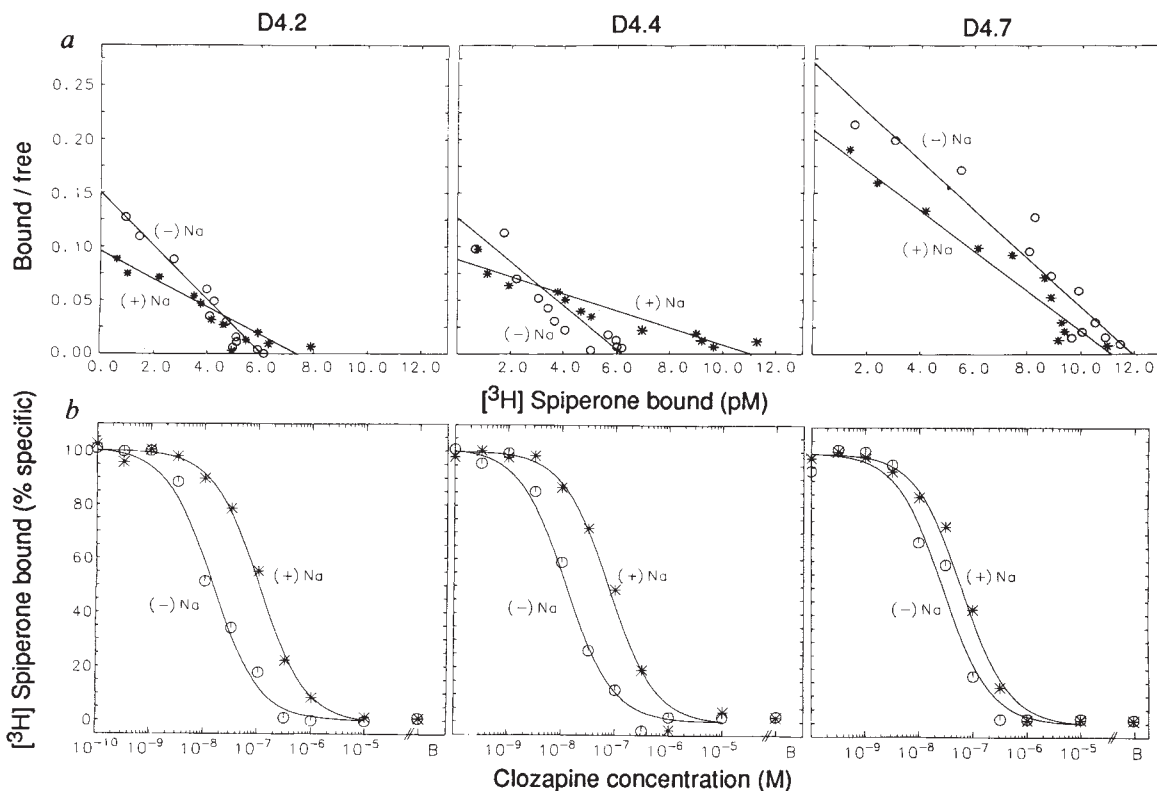
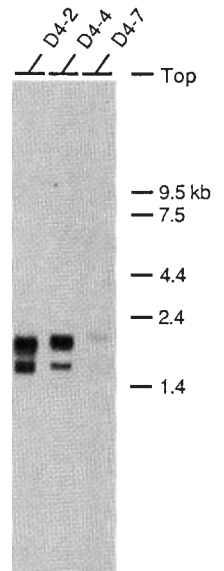


FIG. 4 *a*, Scatchard analysis of the saturation isotherms for [³H]spiperone binding to membranes prepared from COS-7 cells transiently transfected with pCD-D4.2 (D4.2), pCD-D4.4 (D4.4) and pCD-D4.7 (D4.7). *b*, Clozapine competition of [³H]spiperone binding for the three allelic forms of the D4 receptor. The presence of sodium chloride (plus Na⁺ or minus Na⁺) is indicated. To determine accurately differences between the three receptor isoforms, the three clones were analysed simultaneously in four independent experiments for each ligand. Dissociation constants for [³H]spiperone binding (mean $\pm$ s.e. in pM): D4.2 (minus NaCl) 46 $\pm$ 14, (plus NaCl) 81 $\pm$ 18; D4.4 (minus NaCl) 54 $\pm$ 11, (plus NaCl) 100 $\pm$ 23; D4.7 (minus NaCl) 66 $\pm$ 34, (plus NaCl) 67 $\pm$ 24. Dissociation constants for inhibition of [³H]spiperone binding by clozapine (mean $\pm$ s.e. in nM): D4.2 (minus NaCl) 4.2 $\pm$ 0.6, (plus NaCl) 25.3 $\pm$ 3.1; D4.4 (minus NaCl) 3.5 $\pm$ 1.5, (plus NaCl) 28.1 $\pm$ 1.0; D4.7 (minus NaCl) 14.3 $\pm$ 7.5, (plus NaCl) 21.3 $\pm$ 5.3. There is a significant difference in

NaCl sensitivity (ANOVA; F :50.01; P <0.0001) for D4.2 and D4.4 (Student's t test, P <0.0005) but not for D4.7 (t -value 1.3). In the presence of NaCl the three receptors do not show a significant difference (ANOVA; F :2.67; P =0.1231), whereas the difference is revealed in the absence of NaCl (ANOVA; F :5.65; P =0.0257) in which D4.2 and D4.4 are significantly different from D4.7 (Duncan's Range test, α =0.05).

METHODS. For pharmacological analysis¹, COS-7 cells (transiently transfected with one of the three forms of the D4 receptor; see Fig. 3) were homogenized (Polytron, 5 s, setting 5) in 50 mM Tris-HCl (pH 7.4 at 4 °C) buffer containing 5 mM EDTA, 1.5 mM CaCl₂, 5 mM KCl with or without 120 mM NaCl. Specific [³H]spiperone binding was defined as that inhibited by 10 μ M S-sulpiride and/or 30 μ M dopamine, and analysed by the non-linear least-square curve-fitting program LIGAND.

was present. This confirmed the postulated introns 1 and 2 (Fig. 1), and further argues that the repeated region in the putative third cytoplasmic loop is not part of an intron (formerly indicated as intron 3; ref. 1).

In addition, we reconstructed full-length cDNA clones for the variant forms into the expression vector pCD-PS (Fig. 3). Transient expression of the three forms in COS-7 cells and subsequent characterization of the D4 receptor RNA expressed in these cells demonstrated the expected size and size differences between the three forms, indicating that none of the expressed D4 receptor RNAs are further processed (Fig. 3). Furthermore, two bands were observed that correlate with the use of either the endogenous D4 receptor or the SV40 (vector derived) polyadenylation site (Fig. 3). These observations provide additional evidence that the variable sequence is not part of an intron. Furthermore, it indicates that in the transient expression system the expression of the three different clones would result in the formation of three structurally different receptors.

The amino-acid sequences deduced from the different genomic and cDNA clones imply the existence of three different forms of the D4 receptor with a variably sized putative third cytoplasmic loop. Interestingly, this region of the human D4 receptor is not found in the rat homologue of the D4 receptor, making this variation specific to humans (data not shown).

Pharmacological analysis showed that all three variants displayed saturable [³H]spiperone binding (300–1,000 fmol mg⁻¹) with similar dissociation constants (K_d) in the absence of sodium chloride ($K_d \sim 55$ pM; Fig. 4a). But in the presence of 120 mM sodium chloride, the dissociation constants roughly doubled for D4.2 and D4.4 but not for D4.7 (Fig. 4a).

Agonists and antagonists inhibited [³H]spiperone binding (in the presence of sodium chloride) to these different D4 receptor variants in a concentration-dependent manner (dopamine, dissociation constant (K_i) for high-affinity state ~ 25 nM, K_i in presence of guanylimidodiphosphate ~ 460 nM; bromocriptine, $K_i \sim 272$ nM; haloperidol, $K_i \sim 2.5$ nM; raclopride, $K_i \sim 1.6$ μ M; clozapine, $K_i \sim 25$ nM). Furthermore, all three variants exhibited a guanine nucleotide-sensitive high-affinity form of the receptor on competition with dopamine, suggesting that all these variants can functionally couple to G proteins.

Clozapine competition of [³H]spiperone binding showed that D4.2 and D4.4 had significantly (6–8-fold) lower dissociation constants for clozapine in the absence of sodium chloride in the binding assay than they had in its presence. But D4.7 did not exhibit the same sodium chloride sensitivity. The absence of sodium chloride clearly revealed the different pharmacological characteristics between the two shorter variants and D4.7 (Fig. 4). The sodium chloride-mediated effects for clozapine on the D4 variants were not modulated by guanine nucleotides.

The identification of three dopamine D4 receptor variants in the human population is the first example of a polymorphic variation for catecholamine receptors. The fact that the polymorphism is in the domain of the third cytoplasmic loop suggests the possibility of differences in G-protein interaction among the different forms. Polymorphisms due to variable length of repeated sequences in genes have been implicated in diseases such as X-linked spinal and bulbar atrophy¹¹, fragile X-syndrome¹² and myotonic dystrophy¹³. The D4 receptor gene has been localized to chromosome 11p15.5 distal from the Harvey *ras-1* locus¹⁴, a region that has been implicated in bipolar affective disorder¹⁵ and long QT syndrome¹⁶. The observed dopamine D4 receptor polymorphism may have important implications in susceptibility to neuropsychiatric disorders such as schizophrenia and manic depression, as well as response to drug treatment.

Note added in proof: Genetic typing of an additional 200 individuals revealed examples of two more allelic forms of the D4 gene corresponding in size with a three- and five-fold repeat sequence. □

Received 20 March; accepted 8 May 1992.

1. Van Tol, H. H. M. *et al.* *Nature* **350**, 610–614 (1991).
2. Bunzow, J. R. *et al.* *Nature* **336**, 783–787 (1988).
3. Grandy, D. K. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **86**, 9762–9766 (1989).
4. Dal Toso, R. *et al.* *EMBO J.* **8**, 4025–4034 (1989).
5. Sokoloff, P., Giros, B., Martres, M.-P., Bouthenet, M.-L. & Schwartz, J.-C. *Nature* **347**, 146–151 (1990).
6. Kane, J. *et al.* *Arch. gen. Psychiat.* **45**, 789–796 (1988).
7. Casey, D. E. *Psychopharmacology* **99**, 47–53 (1989).
8. Mount, S. M. *Nucleic Acids Res.* **10**, 461–472 (1982).
9. Jackson, I. J. *Nucleic Acids Res.* **19**, 3795–3798 (1991).
10. Bonner, T. I., Young, A., Brann, M. R. & Buckley, N. J. *Neuron* **1**, 403–410 (1988).
11. La Spada, A. R., Wilson, E. M., Lubahn, D. B., Harding, A. E. & Fischback, K. H. *Nature* **352**, 77–79 (1991).
12. Oberle, I. *et al.* *Science* **252**, 1097–1102 (1991).
13. Brook, J. D. *et al.* *Cell* **68**, 799–808 (1992).
14. Gelernter, J., Kennedy, J. L., Van Tol, H. H. M., Civelli, O. & Kidd, K. K. *Genomics* **13**, 208–210 (1992).
15. Kelsoe, J. R. *et al.* *Nature* **342**, 238–242 (1989).
16. Keating, M. *et al.* *Science* **252**, 704–706 (1991).
17. Chomczynski, P. & Sacchi, N. *Analyt. Biochem.* **162**, 156–159 (1987).
18. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning* (Cold Spring Harbor Laboratory Press, New York, 1989).
19. Gubler, U. & Hoffman, B. J. *Gene* **25**, 263–269 (1983).
20. Budowle, B. & Baechtel, F. S. *Appl. theoret. Electrophoresis* (in the press).
21. McMaster, G. K. & Carmichael, G. G. *Proc. natn. Acad. Sci. U.S.A.* **74**, 4835–4838 (1977).

ACKNOWLEDGEMENTS. We thank O. Schoots, S. van Kats, R. Sunahara, A. Petronis, S. Harding and C. Ulpian for assistance; L. Sunahara for human genomic DNA and H. Elsholtz for discussions. This work was supported by the Medical Research Council of Canada (H.H.M.V.T., H.B.N. and P.S.), the Ontario Mental Health Foundation (P.S.) and the NIH (O.C.), H.H.M.V.T. and H.B.N. are Career Scientists of the Ontario Ministry of Health. The sequences of the D4 variants have been submitted to the GenBank and EMBL data bases.

Objective analysis of the topological organization of the primate cortical visual system

Malcolm P. Young

University Laboratory of Physiology, Oxford University, Parks Road, Oxford OX1 3PT, UK

THE primate cortical visual system is composed of many structurally and functionally distinct areas^{1–3}, each receiving and sending about 10 projections from and to other cortical areas¹. The visual cortex is thus served by many cortico–cortical connections to form a network of considerable complexity. Thus the gross organization of this cortical processing system presents a formidable topological problem: although the spatial position of the areas in the brain is reasonably well established, the gross ‘processing architecture’ defined by the connections, is less well understood. Here I report an optimization approach that gives both qualitative and quantitative insight into the connectional topology of the primate cortical visual system. This approach supports suggestions that the system is divided into a dorsal ‘stream’ and a ventral ‘stream’ with limited cross-talk, that these two streams reconverge in the region of the principal sulcus (area 46) and in the superior temporal polysensory areas, that the system is hierarchically organized, and that the majority of the connections are from ‘nearest-neighbour’ and ‘next-door-but-one’ areas.

The cortical sheet can be divided into different areas according to several different parcellation schemes (see, for example, refs 1, 4, 5). I used the most recent parcellation^{1,6}, and examined a matrix of connections between these areas of the macaque cortical visual system (Table 1). The connection matrix was analysed by an optimization method^{7–9} that yields a configuration of points that correspond to the cortical areas. The derived structure optimally fits the matrix so that the distances between the points of the structure are as close as possible to the reverse rank order of the ‘proximities’ between areas in the matrix in Table 1. Thus in the derived structure the length of known connections is at a minimum given the co-constraint that the length of nonexistent (and presently unreported) connections is at a maximum. Hence points representing areas that have very similar afferent and efferent cortico–cortical connections

TABLE 1 Matrix of connections between areas of macaque visual cortex

[illegible]

The cortical parcellation and connections are exactly as in ref. 1 with the exception of MIP and MDP, which have been excluded because of uncertainty concerning their relation to, or identity with, area 7m of ref. 29. Connections coded as '2' are reciprocal, those coded as '1' are one-way projections (whose directions are indicated in Fig. 1) and those coded '0' are either projections which have been explicitly tested for and found absent or connections which are not presently known. No information about the spatial position of the areas in the brain, the laminar termination patterns of the connections, the continuity or patchiness of the distribution of cells giving rise to a projection or about the relative density of projections is represented in the matrix. The information represented concerns only the existence of a connection between two areas, and is therefore the coarsest and most reliable that can be extracted from neuroanatomical studies. Nonetheless, some entries in the matrix can be expected to change as knowledge of the cortical areas and their connections is refined.

are close together, whereas points representing areas that have very different patterns of connections are far apart. The details of the analysis are indicated in the legend to Fig. 1, which shows the best-fit structure for the connectivity matrix in Table 1. Points representing the areas of the cortical visual system are shown connected by the projections to and from the corresponding areas.

Several organizational features of the connectational topology of the system are immediately apparent. The two dimensions of the structure correspond roughly to the posterior-anterior (left to right in Fig. 1) and to the dorsal-ventral (top to bottom) spatial distribution of the areas in the brain. For example, areas of the caudal superior temporal sulcus and of the posterior parietal cortex appear concentrated in the top part of the diagram, whereas areas of the inferotemporal cortex are located in the lower part. Because no information regarding the spatial position of the areas entered the analysis (only information concerning the areas to which each area is connected) this suggests that the spatial position of an area in the brain is a good predictor of the areas to which it is likely to be connected, and that nearby areas tend to innervate one another (see below).

Beginning at the far left of Fig. 1, where primary visual cortex (V1) is located, visual signals pass to a cluster of prestriate areas. This 'prestriate group' consists of areas V2, V3, VP, V4t, V3A, MT and, perhaps surprisingly, area PIP. Areas V3A and MT are topologically less peripheral than other members of this group, and MT is only distinguished topologically by its one-way projections to frontal cortex area 46 and to the frontal eye fields (FEF). Every area of the 'prestriate group' sends output connections to a further cluster of areas comprising FST, MSTd, MSTl, VIP, PO, LIP and DP. The projections from the 'prestriate group'

to this group appear highly redundant, which might account for the fact that partial damage to these prestriate areas does not seriously disrupt spatial vision¹⁰. Signals from the 'MST/posterior parietal complex' then pass to the FEF, area 7a, the posterior part of the superior temporal polysensory area (STPp), and eventually to area 46 and the anterior STP (STPa).

Moving ventrally from V1, signals are relayed by V4 and VOT into the 'inferotemporal (IT) complex'. As V4 is the principal gateway for signals entering the IT, it would seem unlikely on topological grounds that V4 is involved only in colour vision^{3,11,12}. The 'IT complex' seems to be hierarchically organized, in the sense that more anterior stations are topologically further from the periphery, and is associated with parahippocampal areas TF and TH. The topologically 'higher' areas of IT, where some cells respond with high specificity for particular visual patterns¹³⁻¹⁷ project to area 46 and to STPa.

Connections between the dorsal and ventral streams are much less dense than those within each stream, and opportunities for cross-talk do not exist at every station. Both streams, however, project selectively to area 46 and to STPa. Area 46, for example, receives signals which presumably concern what an object is (from IT), where it is (area 7a, LIP), its movement in visual space (MT, MSTd, MSTl), its colour (V4) and its relation to movements of the eyes (FEF).

I used the mathematical tractability of the structure in Fig. 1 to investigate quantitatively whether the topological organization of the visual cortex reflects the dichotomy, reconvergence, hierarchy and border relations suggested by qualitative inspection. This was accomplished by a regression-like procedure, PROCRUSTES rotation¹⁸⁻²⁰, which compared artificial model configurations, numerically embodying each of the proposed

TABLE 2 Quantitative comparisons between Fig. 1 and numerical models

Model	r^2	Probability
Nearest-neighbour	0.29	$P < 0.0017$
Nearest-neighbour or next-door-but-one	0.32	$P < 0.0017$
Hierarchical	0.30	$P < 0.0017$
Two streams and reconvergence	0.46	$P < 0.0017$
Combined hierarchical, two streams and reconvergence	0.72	$P < 0.0017$

Comparisons were effected by a regression-like statistical procedure, PROCUSTES rotation¹⁸⁻²⁰, which finds the optimal reflection, rotation and scaling of each organizational model (for example, nearest-neighbour) with the structure in Fig. 1. At the optimal comparison, the procedure yields a variance-explained statistic which reflects the goodness-of-fit between the two compared models. The statistical rarity of each comparison was assessed by an approximate randomization test^{20,21}, which repeated the PROCUSTES rotation with the organizational model shuffled randomly on each of 600 iterations. The number of times that the variance-explained statistic was exceeded during these random iterations was divided by the number of iterations to yield a probability that a correspondence as good as the particular comparison could have come about by chance^{20,21}. Each r^2 -value in the table is a variance-explained statistic (correlation coefficients may be derived by taking the square root of each r^2 statistic) given by PROCUSTES. Each of the tested organizational models was statistically significantly related to the structure derived for the cortical visual system at a level which would not be expected more than once in 600 comparisons by chance.

organizational features, against the structure in Fig. 1. The statistical rarity of each comparison was assessed by approximate randomization^{20,21} (see Table 2 legend).

Possible border relations embedded in the organization of the visual system were investigated by constructing matrices analogous to Table 1. The first matrix was derived by scoring hypothetical connections between areas that share a common border as '1' and all other possible connections as '0': the 'nearest-neighbour' model (all connections were assumed to be reciprocal). This nearest-neighbour wiring matrix accounted for 61 out of 301 connections in Table 1 (27%). A second matrix was derived by scoring possible connections as a '1' if the areas share a common border or if they are separated by only one intervening area which abuts both areas: the 'nearest-neighbour or next-door-but-one' model. Of the 301 connections in Table 1, 169 (55%) were connections between nearest-neighbour or next-door-but-one areas. Both these matrices were submitted to the same procedure used to derive the structure in Fig. 1, and the resulting configurations were compared with Fig. 1 by PROCUSTES rotation.

The hierarchical ladder derived from the laminar origin and termination patterns of projections^{1,6} was used to construct a unidimensional hierarchical model, by associating an integer value with each area according to its height above V1 in the Felleman and Van Essen scheme. This was the 'hierarchical' model. To model the dichotomization of the system into a dorsal and ventral stream, and the subsequent reconvergence of these two streams, I assigned each area to one of three categories. Posterior parietal areas, caudal superior temporal areas and areas associated with eye movement (such as FEF) were assigned a '3'. 'Shared' areas, such as V1 and V2 (representing the 'shared' origins of the streams in occipital cortex), and STPa and area 46 (representing the reconvergence of the streams), were assigned a '2'. Areas of the ventral stream were assigned a '1'. This constituted the 'two streams and reconvergence' model. Finally, I derived a 'combined hierarchical, two streams and reconvergence' model by making a two-dimensional configuration in which the hierarchical model was dimension 1 and the two streams model was dimension 2.

Table 2 shows the results of the quantitative comparison of these models with the structure in Fig. 1. All five models were related to the structure derived for the real cortical visual system at a level that would not be expected by chance (less than 1 in 600 probability). The two models that represented border relations between areas explained about the same amount of variability (30%) as the hierarchical model. The model that represented dichotomization and reconvergence, however, explained almost half the variability, whereas the combined model accounted for almost three quarters of the variability in Fig. 1.

These results suggest that, despite the enormous complexity of the cortical visual system, at this gross level it may be organized according to four principles. (1) It is dichotomized into two streams, (2) both streams are hierarchies, (3) the streams reconverge in area 46 and STPa, and (4) neighbouring areas tend to innervate one-another.

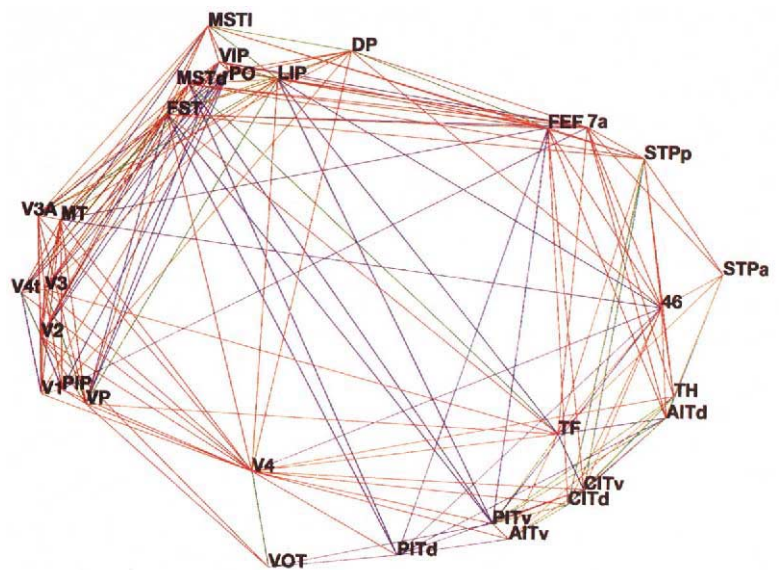
Two of these organizational features derived from topological analysis of the patterns of connections, the two streams and hierarchical features, corroborate organizational principles derived from different information sources, such as lesion studies¹⁰ and from the laminar termination patterns of cortico-cortical projections¹, respectively. The finding that border relations may be present in the wiring pattern of the system might reflect the evolutionary advantage of keeping wiring to a minimum^{22,23}, although the redundancy of connections from the prestriate areas to the areas of the caudal superior temporal sulcus and the posterior parietal cortex suggests that wiring economy is not a strong constraint. Alternatively, the tendency of areas to innervate their neighbours may reflect a parsimonious developmental process.

Some recent models of vision in which the coherence of visual perception is brought about by the synchronization of oscillatory activity in neurons distributed widely in the brain²⁴ have been motivated in part by the apparent lack of a brain structure at which the reconvergence of processed visual information could occur. These results emphasize that there are cortical sites, namely STP^{25,26} and area 46, at which reconvergence may take place. This feature of the primate cortical visual system may allow primates to form coherent representations of their visual environment without the need for stimulus-related oscillatory neuronal activity^{27,28}. □

Received 6 December 1991; accepted 19 May 1992.

- Felleman, D. J. & Van Essen, D. C. *Cereb. Cortex* **1**, 1-47 (1991).
- Peters, A. & Jones, E. G. (eds) *Cerebral Cortex* (Plenum, New York and London, 1985).
- Zeki, S. & Shipp, S. *Nature* **335**, 311-317 (1988).
- Seltzer, B. & Pandya, D. N. *Brain Res.* **149**, 1-24 (1978).
- Bousssaoud, D., Ungerleider, L. G. & Desimone, R. *J. comp. Neurol.* **296**, 462-495 (1990).
- Van Essen, D. C., Felleman, D. J., De Yoe, E. A., Olavarria, J. & Knierim, J. *Cold Spring Harb. Symp. quant. Biol.* **55**, 679-696 (1991).
- Shepard, R. N. *Psychometrika* **27**, 125-140 (1962).
- Young, F. W. *Multidimensional Scaling: History, Theory and Applications* (ed. Hamer, R. M.) (Erlbaum, Hillsdale, New Jersey, 1979).
- Young, F. W. & Harris, D. F. in *SPSS Base System User's Guide* (ed. Norusis, M. J.) 396-461 (SPSS, Chicago, 1990).
- Ungerleider, L. G. & Mishkin, M. in *Analysis of Visual Behavior* (eds Ingle, D. G., Goodale, M. A. & Mansfield, R. J. Q.) 549-586 (MIT, Cambridge, 1982).
- De Yoe, E. A. & Van Essen, D. C. *Trends Neurosci.* **11**, 219-226 (1988).
- Schiller, P. & Lee, K. *Science* **251**, 1251-1253 (1991).
- Miyashita, Y. *Nature* **335**, 817-820 (1988).
- Young, M. P. & Yamane, S. in *Brain Mechanisms of Perception and Memory: from Neuron to Behavior* (eds Ono, T., Squire, L. R., Fukuda, M. & Perrett, D. I.) (Oxford Univ. Press, in the press).
- Young, M. P. & Yamane, S. *Science* **256**, 1327-1331 (1992).
- Tanaka, K., Saito, H., Fukada, Y. & Morioka, M. *J. Neurophysiol.* **66**, 170-189 (1991).
- Gross, C. G., Rocha-Miranda, C. E. & Bender, D. B. *J. Neurophysiol.* **35**, 96-111 (1972).
- Schoneman, P. & Carroll, R. M. *Psychometrika* **31**, 1-10 (1970).
- Gower, J. C. in *Psychometrika* **40**, 33-51 (1975).
- Young, M. P. thesis, St Andrews Univ. (1990).
- Edgington, E. S. *Randomization Tests* (Dekker, New York, 1980).
- Mitchison, G. *Proc. R. Soc. B245*, 151-158 (1991).
- Cowey, A. Q. *J. exp. Psychol.* **31**, 1-17 (1979).
- Crick, F. & Koch, C. *Cold Spring Harb. Symp. quant. Biol.* **55**, 953-962 (1991).
- Baizer, J. S., Ungerleider, L. G. & Desimone, R. *J. Neurosci.* **11**, 168-190 (1991).
- Morel, A. & Bullier, J. *Vis. Neurosci.* **4**, 555-578 (1990).

FIG. 1 The topological organization of the macaque cortical visual system. Reciprocal connections are coloured red, one-way projections going from left to right are coloured blue and one-way projections going from right to left are green. A total of 301 connections is represented, of which 62 are one-way. This non-arbitrary structure is a best-fit representation in 2 dimensions of the connectational topology of this system, in which the positions of areas are specified by their positions being ones that minimize the distance between connected areas and maximize the distance between areas that are not connected. The analysis represents in a spatial framework the organizational structure of the network of cortico-cortical connections between elements of the visual cortex. In detail, the structure was derived by submitting the proximity matrix in Table 1 to non-metric multidimensional scaling⁷⁻⁹, using ALSCAL⁹. Solutions with the level of measurement specified as nominal and ordinal were derived, to assess whether a least-squares categorical transformation was required⁹, but there was no perceptible difference between them. Analogous solutions were derived with different assumptions about the connections which have not been observed (assumed not to exist in Fig. 1). The first alternative solution used a missing data estimation procedure³⁰ to estimate the 'unknown' connections, the second coded 'unknowns' at intermediate values between those for 'known' and 'known not to exist' connections, and the third assumed that all 'unknowns' will be found to exist. These solutions in all cases showed a similar structure to Fig. 1, all explaining more than 76% of its variability, but with areas whose connections are poorly understood (such as VOT and V4t) slightly displaced. Ordinal solutions in one to five dimensions were derived so that solutions with different dimensions could be compared in a 'scree' test. This test showed diminishing returns in numbers of dimensions greater than two.



This structure was accordingly derived with an ordinal level of measurement in two dimensions, and the configuration of points (60 parameters) accounted for 40% of the variability in Table 1 (435 parameters). Despite the results of the 'scree' test, some connections exist between areas which are widely separated in the structure, suggesting that these areas are topologically close in higher dimensions.

27. Young, M. P., Tanaka, K. & Yamane, S. *Soc. Neurosci. Abs.* **17**:73.4 (1991).
 28. Young, M. P., Tanaka, K. & Yamane, S. *J. Neurophysiol.* (in the press).
 29. Cavada, C. & Goldman-Rakic, P. S. *J. comp. Neurol.* **287**, 393-421 (1989).
 30. Norusis, M. J. (ed.) *SPSS Base System User's Guide* (SPSS, Chicago, 1990).

ACKNOWLEDGEMENTS. I thank P. J. Benson, C. Blakemore and K. S. Saleem. This work was supported by a Royal Society Japan Science and Technology Agency research fellowship (during 1990) and by the UK Medical Research Council and the Oxford McDonnell-Pew Centre for Cognitive Neuroscience (during 1991).

Spreading of T-cell autoimmunity to cryptic determinants of an autoantigen

Paul V. Lehmann, Thomas Forsthuber,
Alexander Miller & Eli E. Sercarz*

Department of Microbiology and Molecular Genetics,
University of California at Los Angeles, 405 Hilgard Avenue,
Los Angeles, California 90024, USA

IMMUNIZATION with myelin basic protein (MBP) induces experimental allergic encephalomyelitis (EAE), a prototype of CD4⁺ T-cell mediated autoimmune disease. In rodents, MBP-reactive T-cell clones are specific for a single, dominant determinant on MBP and use a highly restricted number of T-cell receptor genes¹⁻³. Accordingly, EAE has been prevented by various receptor-specific treatments¹⁻⁸, suggesting similar strategies may be useful for therapy of human autoimmune disease. Here we report that in (SJL × B10.PL)F₁ mice, immune dominance of a single determinant, MBP: Ac1-11, is confined to the inductive phase of EAE. In mice with chronic EAE, several additional determinants of MBP in peptides 35-47, 81-100 and 121-140 recall proliferative responses. Most importantly, reactivity to the latter determinants was also detected after induction of EAE with MBP peptide Ac1-11 alone; this demonstrates priming by endogenous MBP determinants. Thus, determinants of MBP that are cryptic⁹⁻¹¹ after primary immunization can become immunogenic in the course of EAE. Diversification of the autoreactive T-cell repertoire due to

'determinant spreading' has major implications for the pathogenesis of, and the therapeutic approach to, T-cell driven autoimmune disease.

We tested the fine specificity of the T-cell response to MBP in (SJL × B10.PL)F₁ mice which are highly susceptible to EAE. Because immunization with autologous MBP is inefficient in priming an autoreactive T cell response we, like others¹⁻⁸, used a xenogeneic protein for priming; guinea-pig MBP shares 91% sequence homology with the murine protein. First, we followed the inductive phase of the immune response in draining lymph nodes after immunization with MBP and pertussis toxin (PTX). Injection of PTX is a prerequisite for induction of EAE in mice, owing to its disruption of the blood-brain barrier¹². At the peak of the local response, there was immune dominance of a single determinant (peptide MBP: Ac1-11) (Fig. 1a). Neither peptide 81-100, the immunodominant MBP determinant in the SJL (H-2^s) strain¹³, nor MBP peptides 35-47 and 121-140, which are both immunogenic and encephalitogenic in the B10.PL (H-2^u) mice^{14,15}, induced a response in T cells from F₁ mice after immunization with MBP (Fig. 1a). Therefore, peptides 35-47, 81-100 and 121-140 behaved as cryptic determinants. The fine specificity of the T-cell response to MBP in the splenocyte population on day 9 was essentially identical (Fig. 1b). As a prototype 'foreign' antigen, we used hen eggwhite lysozyme (HEL) for control immunizations. In HEL/PTX-immunized F₁ mice, peptide HEL: 30-53 was immunodominant (Fig. 1d, e). Importantly, HEL peptides 11-25 and 116-129, which are both dominant in the SJL parental strain (and which are immunogenic when injected as peptides into F₁ mice; P.V.L. and T.F., unpublished observations), behaved as cryptic determinants in the F₁ mice (Fig. 1d, e). Thus, in both the MBP and HEL systems, immune dominance of a single determinant occurred despite the presence of several potential T-cell determinants

* To whom correspondence should be addressed.

on the antigen and in the presence of a variety of restriction elements in the host.

To account for the possibility that the autoreactive repertoire diversifies in chronic EAE¹⁶⁻¹⁸, we defined the fine specificity of the MBP response 40 days after immunization with MBP/PTX. Strong proliferative responses were recalled by MBP peptides Ac1-11, as well as 35-47, 81-100 and 121-140 in the spleen (Fig. 1c). (At this time point, the draining lymph nodes no longer responded.) In contrast, F₁ mice that had been immunized with HEL/PTX, showed no changes with time in the fine specificity of the T-cell response to the 'foreign' antigen, HEL (Fig. 1f). Murine and hen lysozyme share 56% sequence homology. Unlike MBP, there is profound self tolerance established to autologous lysozyme; immunization of mice with HEL does not induce a detectable autoimmune condition. Thus, determinant spreading occurred only with MBP.

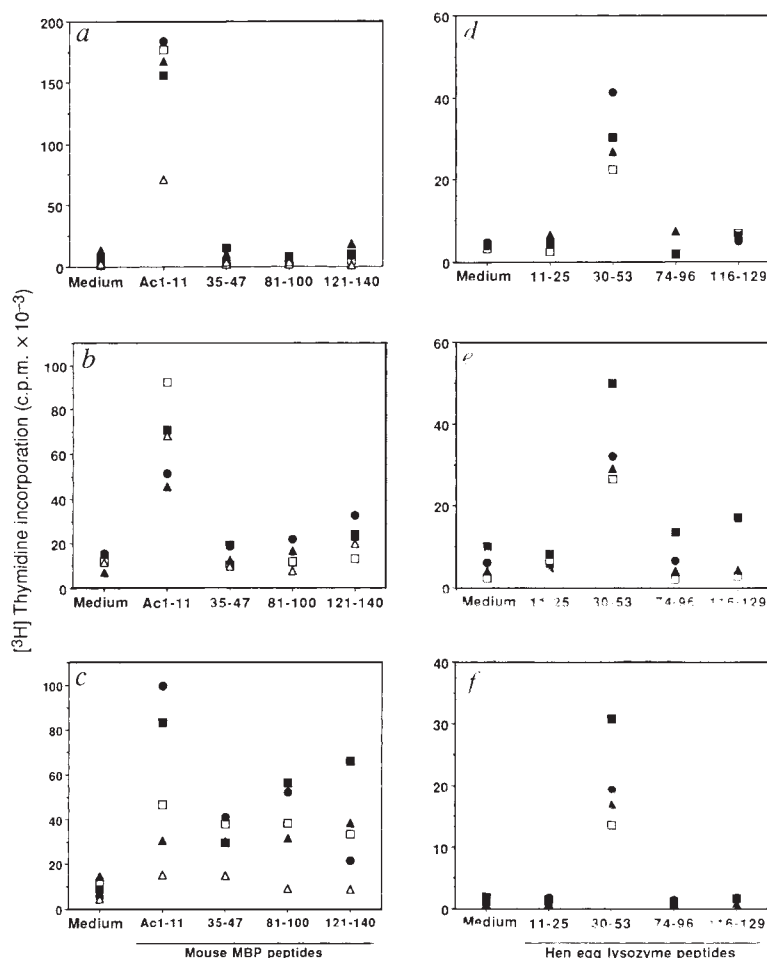
These data show that peptides of MBP which were cryptic during primary immunization behaved as dominant determinants in chronically diseased mice. It is possible that altered presentation of endogenous MBP occurred in the inflamed central nervous system (CNS), resulting in a second wave of priming by endogenous antigen to peptides 35-47, 81-100 and 121-140. To address this possibility, we provoked EAE solely with Ac1-11/PTX and tested the recirculating lymphocyte pool 21 days

later for recall to MBP peptides. Responses to formerly cryptic determinants were now detected (Fig. 2a). We noted the relative dominance of MBP:121-140 (Fig. 2a), suggesting that a hierarchy of responsiveness in the group of formerly cryptic determinants had become established. Because the HEL:30-53/PTX-immunized control group did not reveal MBP reactivity (Fig. 2b), disruption of the blood-brain barrier by PTX¹² did not induce an autoimmune response. Rather, T cell-triggered inflammation in the CNS, including local interferon- γ release, seemed to be required for priming to formerly 'cryptic self'^{11,19}.

Enhanced presentation of microbial antigens at sites of chronic, T-cell mediated inflammation might be critical for overcoming infections. This lymphokine-mediated reaction should assist immune surveillance by inducing a second round of priming to formerly cryptic determinants of the pathogen. In EAE, the same process of determinant recruitment seemed to lead to the broadening of the initially restricted T-cell response to MBP. The previously cryptic MBP determinants are encephalitogenic when injected as peptides^{14,15}, and therefore it can be expected that primed T cells reacting to these determinants also contribute to the pathogenesis of chronic, relapsing EAE. Being directed against multiple determinants of MBP, the autoreactive T-cell repertoire is likely to be heterogenous with respect to T-cell receptor (TCR) gene usage (unless autoreactive

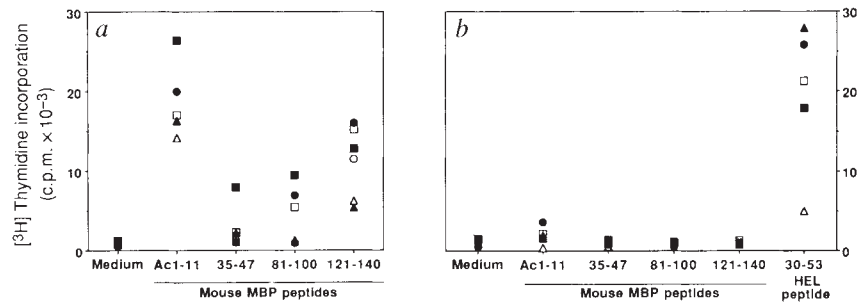
FIG. 1 Dominance of a single antigenic determinant in the early T-cell response to MBP and HEL; diversification of the MBP, but not of the HEL-specific response with time. Groups of 4 to 5 (SJL $\times$ B10.PL)_{F1} mice were immunized with guinea pig MBP and PTX (a, b, c) or HEL and PTX (d, e, f). Nine days later, the proliferative recall response to peptides which correspond to the murine MBP sequence or to HEL was tested as indicated in the regional lymph node (a, d) or in the spleen (b, e). The splenocyte response was also measured 40 days after immunization in F₁ mice that exhibited grade 3-4 severity¹ of EAE (c, f). Different symbols in each panel denote results obtained with individual mice. The data are expressed as arithmetic means of [³H]thymidine incorporation (c.p.m. $\times 10^{-3}$) in triplicate cultures; standard deviation between replicate cultures was $< \pm 10\%$. Each panel shows one representative experiment of the 3 to 4 done.

METHODS. SJL (H-2^b) and B10.PL (H-2^d) mice were obtained from the Jackson Laboratory (Bar Harbor, ME) and (SJL $\times$ B10.PL)_{F1} hybrids were bred at UCLA. Mice of either sex were used at 3 to 5 months of age. The mice were immunized subcutaneously with 100 μ g guinea-pig MBP or HEL in complete Freund's adjuvant (CFA) and 0.1 μ g PTX (List Biological, Campbell, CA) was injected in 0.5 ml saline intravenously (tail vein) 24 and 72 h later. At the time points indicated, draining (popliteal) lymph nodes and spleens of mice were removed and single cell suspensions were prepared. Lymph node cells were cultured at 4×10^5 , and splenocytes were cultured at 8×10^5 , cells per well in 96-well microtitre plates in 200 μ l serum-free medium (HL-1; Ventrex, Portland, ME) supplemented with 2 mM glutamine; peptides were added as indicated at 7 μ M final concentration which was optimal for recall of a response. Proliferation was assayed by addition of 1 μ Ci [³H]thymidine (International Chemical and Nuclear, Irvine, CA) for the last 18 h of a 5 day culture, and incorporation of label was measured by liquid scintillation counting. MBP was isolated from brains of guinea-pigs (Pel-Freez Biologicals, Rogers, AR) as described²². HEL was obtained from Societa Prodotti Antibiotici (Milan, Italy) and further purified by elution from a BioRex-70 column (Bio-Rad). MBP peptides were synthesized by S. Horvath, except for Ac1-11 which was prepared in our laboratory; HEL peptides were made by J. Young, M. McMillan and L. Williams using a solid-phase technique²³. Peptides were purified by high-performance liquid chromatography (HPLC) on a preparative C₈ column with a solvent system of 0.1% trifluoroacetic acid and an increasing gradient of acetonitrile (the HEL:74-96 peptide preparation gave a single major peak on analytical HPLC and, because of its poor solubility, was not purified further). Peptide identity was confirmed by amino-acid composition analysis. The MBP peptides correspond to the



murine protein and the sequences are as follows (single-letter amino-acid code): Ac1-11, ACASQKRPSQSRK; 35-47, ILDSIGRFFSGDR; 81-100, NPVVHFFKNIVTPRTPPPSQ; 121-140, GFGYGRASDYKSAHKGFKG. The HEL peptide sequences are: 11-25, AMKRHGLDNYRGYSL; 30-53, CAAK-FESNFNTQATNRNTDGTSDY; 74-96, NLNIPCSALLSSDITASVNCAL; 116-129, KGTDVQAWIRGCR.

FIG. 2 Detection of T cells primed to MBP determinants 35–47, 81–100 and 121–140 in spleens of Ac1–11-immunized mice. (SJL × B10.PL)_{F1} mice were immunized with MBP:Ac1–11 and PTX (a) or HEL:30–53 and PTX (b) and splenocytes of both groups were assayed 21 days later for proliferation to the indicated peptides. The MBP:Ac1–11 immunized mice had developed EAE of stage 1–2. The data shown are arithmetic means of individual responses (each mouse is represented by a different symbol). One experiment is shown of the three done; 17 mice have been tested individually in three experimental groups with similar results. METHODS. Mice were immunized with 7 nmol of the indicated peptide in CFA subcutaneously; 0.1 µg PTX was also injected intravenously 24 and 72 h later. The splenocyte proliferation assay was done as described in the legend to Fig. 1.



T cells with unrelated specificities share TCR genes²⁰). By interfering with the initial dominant response, anti-clonotypic treatment was shown to prevent acute EAE^{1–6}. Treatment in the first 3 days after onset of the clinical disease seems to be curative^{2,7,8}. In these experiments, we presume that inactivation of the first wave of peripherally primed, still clonally restricted effector cells also abrogated the inflammatory reaction in the CNS, a prerequisite for second-wave priming (Fig. 2). But our data raise the possibility that after determinant spreading has occurred, anti-clonotypic therapy may no longer be effective in interfering with the diversified autoreactive T-cell pool in the more advanced disease.

Considering EAE as a model for T-cell mediated autoimmune disease in general, and multiple sclerosis in particular, our data predict that as human autoimmunity develops, priming to additional determinants on self antigens might occur. This amplification and broadening of the autoimmune response may be critical in determining the progressive course of human autoimmune disease²¹ as well as altering the therapeutic approach to it. Therefore, prevention of second-wave priming to autoantigens might be a prerequisite for successful therapy. □

Received 20 December 1991; accepted 8 May 1992.

- Urban, J. L. *et al. Cell* **54**, 577–592 (1988).
- Acha-Orbea, H. *et al. Cell* **54**, 263–273 (1988).
- Ohashi, M. & Heber-Katz, E. *J. exp. Med.* **168**, 2153–2164 (1988).
- Ben-Nun, A., Wekerle, H. & Cohen, I. R. *Nature* **292**, 60–61 (1981).
- Vandenbark, A. A., Hashim, G. & Offner, H. *Nature* **341**, 541–544 (1989).
- Howell, M. D. *et al. Science* **246**, 668–670 (1989).
- Offner, H., Hashim, G. A. & Vandenbark, A. *Science* **251**, 430–432 (1991).
- Zaller, D. M., Osman, G., Kanagawa, O. & Hood, L. *J. exp. Med.* **171**, 1943–1955 (1990).
- Gammon, G. & Sercarz, E. E. *Nature* **342**, 183–185 (1989).
- Benichou, G. *et al. J. exp. Med.* **172**, 1341–1346 (1990).
- Gammon, G., Sercarz, E. E. & Benichou, G. *Immun. Today* **12**, 193–195 (1991).
- Linthicum, D. S. & Frelinger, J. A. *J. exp. Med.* **155**, 31–40 (1982).
- Sakai, K. *et al. J. Neuroimmun.* **19**, 21–32 (1988).
- Hood, L. *et al. Cold Spring Harb. Symp. quant. molec. Biol.* **54**, 859–874 (1989).
- Zamvil, S. S. *et al. J. exp. Med.* **168**, 1181–1186 (1988).
- Perry, L. L. & Barzaga, M. E. *J. Immun.* **138**, 1434–1441 (1987).
- Perry, L. L., Barzaga-Gilbert, E. & Trotter, J. L. *J. Neuroimmun.* **33**, 7–15 (1991).
- McCarron, R. M., Fallis, R. J. & McFarlin, D. E. *J. Neuroimmun.* **29**, 73–79 (1990).
- Sarvetnick, N. *Nature* **346**, 844–847 (1990).
- Heber-Katz, E. & Acha-Orbea, H. *Immun. Today* **10**, 164–169 (1989).
- Fritz, R. B. & McFarlin, D. E. in *Chem. Immun.* **46**, 101–125 (1989).
- Diebler, G. E., Martenson, R. E. & Kies, M. W. *Prep. Biochem.* **2**, 139–165 (1972).
- Kent, S. & Clark-Lewis, I. in *Synthetic Peptides in Biology and Medicine* (eds Alitalo, K., Partonen, P. & Vahen, A.) 29–42 (Elsevier, Amsterdam, 1985).

ACKNOWLEDGEMENTS. We thank Z. A. Nagy, M. Kronenberg and C. Uittenbogaart for discussions and critical reading of the manuscript. S. Horvath (Caltech), J. Young (UCLA), M. McMillan and L. Williams (USC) for synthesizing the peptides, and R. Trezza for technical assistance. This work was supported by grants from the NIH and the Multiple Sclerosis Society. P.V.L. and T.F. are fellows of the Deutsche Forschungsgemeinschaft.

Platelet-derived growth factor stimulates synthesis of PtdIns(3,4,5)P₃ by activating a PtdIns(4,5)P₂ 3-OH kinase

P. T. Hawkins, T. R. Jackson & L. R. Stephens

Biochemistry Department, Institute of Animal Physiology and Genetics Research, AFRC, Babraham, Cambridge CB2 4AT, UK

ALTHOUGH the hormone-stimulated synthesis of 3-phosphorylated inositol lipids is known to form an intracellular signalling system^{1–5}, there is no consensus on the crucial receptor-regulated event in this pathway and it is still not clear which of the intermediates represent potential output signals. We show here that the key step in the synthesis of 3-phosphorylated inositol lipids in 3T3 cells stimulated by platelet-derived growth factor is the activation of a phosphatidylinositol(4,5)-bisphosphate (3)-hydroxy (PtdIns(4,5)P₂ 3-OH) kinase. A similar conclusion has been applied to explain the actions of formyl-Met-Leu-Phe on neutrophils⁶, and it may be that receptors that couple through intrinsic tyrosine kinases or through G proteins stimulate the same step in 3-phosphorylated inositol lipid metabolism. The close parallel between these two mechanisms for the activation of PtdIns(4,5)P₂ 3-OH kinase and those described for the activation

of another key signalling enzyme, phospholipase C (ref. 7), focuses attention on the product of the PtdIns(4,5)P₂ 3-OH kinase, PtdIns(3,4,5)P₃, as a possible new second messenger.

We have investigated the pathway for the formation of 3-phosphorylated inositol lipids in platelet-derived growth factor (PDGF)-stimulated 3T3 cells by using the same two whole-cell techniques that we applied to fMet-Leu-Phe (FMLP)-stimulated neutrophils: detailed measurement of the changes in concentration of the polyphosphoinositides, and an analysis of the relative rates at which ³²P-labelled inorganic phosphate ([³²P]P_i) added to the medium of the cells is incorporated into individual phosphate moieties of the polyphosphoinositides (see Table 1 legend, and refs 6 and 8 for a detailed consideration of this strategy).

For the first approach, Swiss 3T3 cells were labelled with [³²P]P_i for 70 min, conditions that resulted in changes in the amount of radioactivity in each of the polyphosphoinositides being very largely a function of changes in their concentration (see Table 1 legend, protocol A). Under these conditions, PDGF caused a rapid and sustained increase in the levels of PtdIns(3,4,5)P₃ and PtdIns(3,4)P₂, and no significant change in PtdIns(3)P over the first five minutes of stimulation (Fig. 1; ref. 9). Further, the data convincingly demonstrate a substantial lag in the PDGF-stimulated production of PtdIns(3,4)P₂ relative to PtdIns(3,4,5)P₃. This lag is a powerful argument in favour of PtdIns(3,4,5)P₃ being the primary output signal from this pathway and also places significant constraints on the origins of PtdIns(3,4)P₂ (see later).

TABLE 1 Analysis of the distribution of ^{32}P among the individual phosphate groups of polyphosphoinositides extracted from PDGF-stimulated 3T3 cells

Experimental protocol	Phospholipid	Total ^{32}P d.p.m. in analysis	% Distribution of ^{32}P among individual phosphates			
			1 (Phosphodiester)	3	4	5
A	PtdIns(4,5) P_2	247,622	1.7		48.3	50.0
B	PtdIns(4,5) P_2	908,921	0.2		45.8	54.0
	PtdIns(3,4) P_2	4,002	0.5	53	46.5	
	PtdIns(3,4,5) P_3	13,632	0.45	41.3	26.0	32.0

Swiss 3T3 cells were grown to confluence and serum-starved for 16 h. Cells were grown in small glass bottles and handled as described in Fig. 1 legend. Two experiments were run with different ^{32}P -labelling regimes: (A) essentially as described for Fig. 1, or (B) cells were labelled with $10.5 \text{ mCi } [^{32}\text{P}]\text{P}_i \text{ ml}^{-1}$ for 30 min, then stimulated with PDGF (30 ng ml^{-1}) for 35 s, while still in the presence of the added $[^{32}\text{P}]\text{P}_i$. Incubations were quenched and the lipids extracted as described in the legend to Fig. 1. Dried lipid extracts derived from cells processed by both protocols A and B were deacylated and resolved by anion exchange HPLC as described⁶. The ^{32}P -labelled peaks eluting at the times expected for GroPIns(3,4) P_2 (glycerophospho-inositol(3,4)-bisphosphate), GroPIns(4,5) P_2 and GroPIns(3,4,5) P_3 were pooled, desalted and the distribution of ^{32}P among the individual phosphates was determined as described⁶, the only differences being that the ^3H -labelled analogues used to calibrate the analysis ($[^3\text{H}]\text{Ins}(1,3,4)\text{P}_3$, $[^3\text{H}]\text{Ins}(1,4,5)\text{P}_3$ and $[^3\text{H}]\text{Ins}(1,3,4,5)\text{P}_4$) were prepared as described⁸, then mixed with the deglycerated, ^{32}P -labelled, 3T3 cell-derived samples before the series of dephosphorylations that form the basis of these experiments. This meant that in this analysis, the alkaline phosphatase-catalysed dephosphorylations used to determine the ^{32}P content of the phosphodiester phosphates of the original glycerophosphoinositol phosphates were standardized by calculating the percentage of the total ^{32}P recovered from the assays that was in $[^{32}\text{P}]\text{P}_i$. Detailed considerations of the overall validity of this form of analysis have been published elsewhere^{6,8}. Under the pre-equilibrium conditions of protocol B, the relative ^{32}P content of a particular phosphate allows it to be ranked in terms of its speed of labelling and this in turn defines the sequences of phosphorylation reactions that lead to the formation of that compound, so the most rapidly labelled phosphate was added last, the most slowly was added first, and so on. Thus, under the conditions of protocol B the 5-phosphate of PtdIns(4,5) P_2 was slightly more radioactive than the 4-phosphate, consistent with both its expected route of synthesis from PtdIns4P and also the rapid rate of metabolism of these two compounds. In contrast, under the more extensive labelling conditions of protocol A, there was a near-equilibrium distribution of $[^{32}\text{P}]\text{P}_i$ between the 4- and 5-phosphates of PtdIns(4,5) P_2 , suggesting that both monoester phosphates had reached isotopic equilibrium with the γ -phosphate of ATP. This result in turn suggests that under the labelling conditions of protocol A (used in Fig. 1), changes in the amounts of radioactivity in the polyphosphoinositides are likely to reflect closely changes in their concentrations.

For the second approach, we labelled the cells for 30 min with $[^{32}\text{P}]\text{P}_i$ and then stimulated them with a maximal dose of PDGF for only 35 seconds to allow us to establish the relative rates of labelling of the individual phosphates of these lipids during stimulation. This design ensured that the processes leading to incorporation of ^{32}P into the individual phosphates were far from steady-state, a condition upon which this whole

approach hinges (see Table 1 legend, and refs 6 and 8). The data presented in Table 1 (protocol B) show that under these conditions the 3-phosphate of PtdIns(3,4,5) P_3 was the most radioactive and therefore that this lipid is synthesized by a PtdIns(4,5) P_2 3-OH kinase; this conclusion is supported by the similarity in the distribution of $[^{32}\text{P}]\text{P}_i$ among the 1-4- and 5-phosphates in this precursor/product pair. As the levels of

FIG. 1 Time course of PDGF-stimulated formation of $[^{32}\text{P}]\text{PtdIns}(3,4,5)\text{P}_3$ (circles) and $[^{32}\text{P}]\text{PtdIns}(3,4)\text{P}_2$ (filled circles) in ^{32}P -labelled 3T3 cells. The data shown define the quantity (in d.p.m.) of $[^{32}\text{P}]\text{PtdIns}(3,4)\text{P}_2$ and $[^{32}\text{P}]\text{PtdIns}(3,4,5)\text{P}_3$ derived from individual bottles of 3T3 cells which had been incubated for various times with either vehicle or PDGF (added at $t=0$). Data are drawn from 3 independent experiments and are for the PDGF-stimulated samples after subtraction of control values (control values did not change significantly with time throughout the time course). The combined control values for each lipid were ($\bar{x} \pm \text{s.e.m.}$, $N=15$) PtdIns3P, 906 ± 40 ; PtdIns(4)P, $16,231 \pm 626$; PtdIns(3,4) P_2 , 110 ± 12 ; PtdIns(4,5) P_2 , $67,097 \pm 1,844$; PtdIns(3,4,5) P_3 , 121 ± 8 . The Gompertz curves shown ($y=c \exp(-\exp[-b(t-m)])$), constrained to have a lower asymptote of zero were fitted to the data by nonlinear regression analysis using the Genstat Package¹⁹. Data were significantly ($P < 0.001$) better fitted by two curves displaying a relative lag (b , m and c values different for each compound) than a single curve of the same type (b and m values same for each compound, but allowing the linear scaling factor c to be different).

METHODS. Swiss 3T3 cells were grown to confluence over six days^{9,20} in small glass bottles (Wheaton 'shorty' vials, $19 \text{ mm} \times 40 \text{ mm}$) and then serum-starved for 16 h before the experiment. Each bottle was then removed from the incubator, washed with HEPES-buffered DMEM (112.5 mM NaCl , 5.37 mM KCl , 0.81 mM MgSO_4 , 1.8 mM CaCl_2 , 25 mM glucose , 1 mM NaHCO_3 , 25 mM HEPES , pH 7.4, at 37°C , 0.1% fatty-acid free BSA, $1 \times \text{DMEM}$ amino acids and vitamins) and then incubated with 0.4 ml HEPES-buffered DMEM containing $0.25 \text{ mCi } [^{32}\text{P}]\text{P}_i$ for 70 min at 37°C . Each bottle was then washed twice and incubated at 37°C in a total volume of 0.4 ml HEPES-buffered DMEM either with or without PDGF (human recombinant form BB (gift from K. D. Brown) used at a final concentration of 30 ng ml^{-1}). Incubations were terminated by addition of 1.5 ml $\text{CHCl}_3/\text{MeOH}$ ($1:2 \text{ v/v}$). Lipids were extracted using a Folch distribution of organic and aqueous phases, with 1.0 M HCl , 25 mM EDTA , $5 \text{ mM tetrabutylammonium sulphate}$ included in the aqueous phase⁹. Lipids were deacylated with methylamine and the glycerophosphoinositol esters separated by HPLC using a Partisphere SAX column as described⁶. The identities of the appropriate radioactive compounds obtained using this strategy have been rigorously characterized⁶.

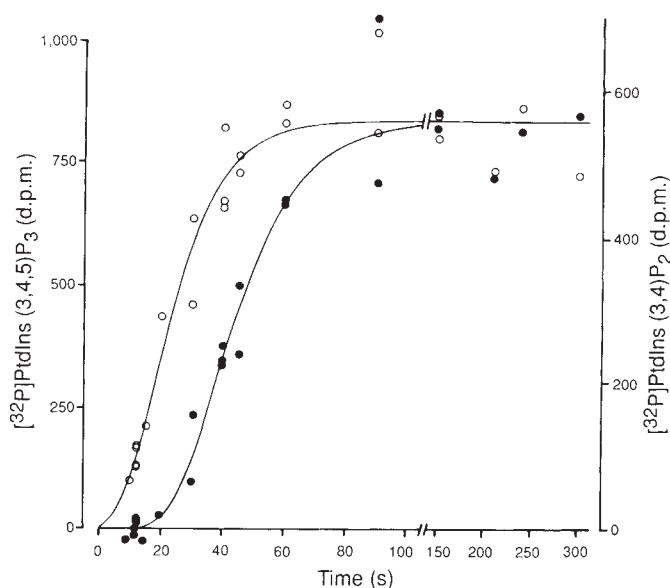


TABLE 2 Formation of PtdIns(3,4)P₂ *in vitro*

Source of tissue	Dephosphorylation of PtdIns([³² P]-3,4,5)P ₃		Percentage of total ³² P-labelled products formed	
	Substrate degraded	Percentage per min per (mg per ml)*	PtdIns(3,4)P ₂	P _i
Lysate (0.438 mg ml ⁻¹ protein)	7.51 ± 0.24	11.3 ± 0.38	24.8 ± 2.5	75.2 ± 0.67
Soluble fraction (0.225 mg ml ⁻¹ protein)	5.62 ± 0.16	16.6 ± 0.46	20.4 ± 0.8	79.5 ± 2.67
Particulate fraction (0.225 mg ml ⁻¹ protein)	1.64 ± 0.13	4.76 ± 0.36	87.2 ± 7.9	12.8 ± 0.12

3T3 cell lysates, soluble and particulate fractions were prepared essentially as described in ref. 6 and 10 µl aliquots were added to prewarmed tubes containing 30 µl assay buffer (to give final concentrations of 1 mM ATP, 4 mM MgCl₂, 1 mM EGTA, 50 mM HEPES, pH 7.2, at 37 °C, 70 mM potassium glutamate, 25 µM EDTA, sufficient calcium to yield a final free calcium of 100 nM, a physiological mix of phospholipids⁶, including 30 or 12 µM PtdIns4P with or without 712 × 10³ d.p.m. ml⁻¹ [³H]PtdIns4P (1 Ci mmol⁻¹; Amersham), 30 µM PtdIns(4,5)P₂ with or without 712 × 10³ d.p.m. ml⁻¹ [³H]PtdIns(4,5)P₂ (1 Ci mmol⁻¹; Amersham) and 0.1 nM PtdIns(3,4,5)P₃ containing 712 × 10³ d.p.m. ml⁻¹ PtdIns(³²P-3,4,5)P₃ (prepared as described in ref. 6; 3,000 Ci mmol⁻¹). Under the conditions of these assays, the relevant activities operated linearly with respect to time and protein concentration (data not shown) and the only detectable ³²P-labelled products were [³²P]Pi and [³²P]PtdIns(3,4)P₂. Data shown are means ± s.e.m. (n = 3). A minimum of 25% of the PtdInsP₃ hydrolysed in 3T3 cell lysates formed PtdIns(3,4)P₂ (this is a minimum estimate because it is possible that the [³²P]Pi released is derived from PtdIns(3,4)P₂ formed during the assays). About 55% of the PtdIns(3,4,5)P₃ 5-phosphatase present in the 3T3 lysates was recovered in particulate fractions. When assayed under identical conditions 3T3 cell lysates converted [³H]PtdIns(4,5)P₂ to [³H]PtdIns(3,4,5)P₃ at a rate of 2.48 min⁻¹ mg ml⁻¹ (0.84 nmol ml⁻¹ min⁻¹ mg ml⁻¹), an analogous PtdIns(4)P 3-OH kinase activity could not be detected in parallel assays (limits of detection were 0.2–0.3% min⁻¹ mg ml⁻¹). It is difficult to extrapolate these *in vitro* results to the intact cell (because of profound differences in substrate presentation) but they do suggest that a substantial proportion of the PtdIns(3,4)P₂ formed in PDGF-stimulated 3T3 cells may be derived from PtdIns(3,4,5)P₃.

* For substrate degraded, 100% is 100 f mol ml⁻¹.

[³²P]PtdIns(3,4,5)P₃ were still rising roughly linearly after 35 seconds stimulation while those of [³²P]PtdIns(4,5)P₂ were essentially unchanged, and as PDGF stimulates the initial rate of incorporation of ³²P from [^γ-³²P]ATP into PtdIns(3,4,5)P₃ in streptolysin O-permeabilized 3T3 cells⁹, these results strongly suggest that PDGF activates a PtdIns(4,5)P₂ 3-OH kinase (as opposed to increasing the supply of [³²P]PtdIns(4,5)P₂, and hence production of [³²P]PtdIns(3,4,5)P₃, by simple mass action, or by inhibiting an enzyme that degrades [³²P]PtdIns(3,4,5)P₃).

The 3-phosphate of PtdIns(3,4)P₂ was more radioactive than the 4-phosphate (Table 1, protocol B). This result, together with those already described showing that the 3-phosphate is also the most radioactive in PtdIns(3,4,5)P₃, restricts the possible known precursors of PtdIns(3,4)P₂ to either PtdIns(4)P (by direct 3-phosphorylation) or PtdIns(3,4,5)P₃ (by means of a 5-phosphomonoesterase). Both the PtdIns(4)P 3-OH kinase¹⁰ and PtdIns(3,4,5)P₃ 5-phosphatase⁶ (Table 2) activities demanded by these two alternatives have been detected in various cell lysates, but we are unable to predict confidently their relative magnitudes in the intact cell (Table 2 legend). The significant lag that we observed in the PDGF-stimulated accumulation of PtdIns(3,4)P₂ relative to PtdIns(3,4,5)P₃ (Fig. 1) demands that for PtdIns(4)P to be the precursor of PtdIns(3,4)P₂ then either (1) the substrate specificity and/or supply of the putative PtdIns(4)P 3-OH kinase must change during the first seconds of stimulation (an effect that might be produced if it was internalized along with the PDGF receptor to an endosomal membrane relatively deficient in PtdIns(4,5)P₂), or (2) the putative PtdIns(4)P 3-OH kinase is distinct from, and shows different kinetics of activation to, the enzyme that generates

PtdIns(3,4,5)P₃. The alternative explanation, that PtdIns(3,4)P₂ is derived by dephosphorylation of PtdIns(3,4,5)P₃, assumes that the sources of these lipids are different. Hence the presence of a lag in PtdIns(3,4)P₂ accumulation becomes a natural consequence of the additional steps involved in its formation. We favour this latter explanation on the grounds that it is the most simple, requires the minimum number of qualifications and 'accounts' for the degradative metabolism (that is, turnover) of the polyphosphoinositides without the need to postulate the existence of further intermediates, metabolic pools or enzymes.

The results described here for PDGF-stimulated Swiss 3T3 cells are qualitatively identical to those we recently described for neutrophils stimulated by the formyl-peptide fMet-Leu-Phe and suggest that the primary action of both PDGF and the formyl-peptide is to stimulate the synthesis of PtdIns(3,4,5)P₃ by activating a PtdIns(4,5)P₂-selective 3-OH kinase. This conclusion disagrees with previous metabolic schemes proposed to account for agonist-induced changes in 3-phosphorylated inositol lipid metabolism^{1,10–12}. If our interpretation of the results is correct, then this pathway meets one of the critical criteria for its consideration as a signalling system (the origin of this idea is reviewed in ref. 1): that of possessing a common, pivotal enzyme through which various ligands funnel their signals. Natural conclusions to this line of reasoning are that the inositol lipid 3-OH kinase that has been shown physically to associate with activated tyrosine kinase-containing receptors (see for example, refs 13–18) functions as a PtdIns(4,5)P₂ 3-OH kinase in the intact cell, and moreover that PtdIns(3,4,5)P₃ is the primary candidate for the role of output signal from this putative signal transduction cascade. □

Received 9 April; accepted 13 May 1992.

1. Carpenter, C. L. & Cantley, L. C. *Biochemistry* **29**, 11147–11156 (1990).
2. Sugimoto, Y., Whitman, M., Cantley, L. & Erikson, R. L. *Proc. natn. Acad. Sci. U.S.A.* **81**, 2117–2121 (1984).
3. Macara, I. G., Marinatti, G. V. & Balduzzi, P. C. *Proc. natn. Acad. Sci. U.S.A.* **81**, 2728–2732 (1984).
4. Whitman, M., Downes, C. P., Keeler, M., Keller, T. & Cantley, L. *Nature* **332**, 644–646 (1988).
5. Traynor-Kaplan, A. E., Harris, A. L., Thompson, B. L., Taylor, P. & Sklar, L. A. *Nature* **334**, 353–356 (1988).
6. Stephens, L. R., Hughes, K. T. & Irvine, R. F. *Nature* **351**, 33–39 (1991).
7. Rhee, S. G. *Trends bio. Sci.* **16**, 297–301 (1991).
8. Stephens, L. R. & Downes, C. P. *Biochem. J.* **265**, 435–452 (1990).
9. Jackson, T. R., Stephens, L. R. & Hawkins, P. T. *J. biol. Chem.* (in the press).
10. Auger, K. R., Serunian, L. A., Soltoff, S. P., Libby, P. & Cantley, L. C. *Cell* **57**, 167–175 (1989).
11. Cunningham, T. W. *et al. J. biol. Chem.* **265**, 21676–21683 (1990).
12. Cunningham, T. W. & Majerus, P. W. *Biochem. biophys. Res. Commun.* **175**, 568–576 (1991).

13. Whitman, M., Kaplan, D., Roberts, T. & Cantley, L. *Biochem. J.* **247**, 165–174 (1987).
14. Kazlauskas, A. & Cooper, J. A. *EMBO J.* **9**, 3279–3286 (1990).
15. Escobedo, J. A. *et al. Cell* **65**, 75–82 (1991).
16. Skolnik, E. Y. *et al. Cell* **65**, 83–90 (1991).
17. Otsu, M. *et al. Cell* **65**, 91–104 (1991).
18. Shibasaki, F., Homma, Y. & Takenawa, T. *J. biol. Chem.* **266**, 8108–8114 (1991).
19. Payne, R. W. *et al. Genstat 5 Reference Manual* (Clarendon, Oxford University Press, UK, 1987).
20. Brown, K. D. & Blakeley, D. M. *Biochem. J.* **212**, 465–472 (1983).

ACKNOWLEDGEMENTS. We thank D. Brown for statistical analysis of the time course data; R. Tregear, N. Divecha, T. Corps, R. F. Irvine, A. Eguinoa and D. Poyner for discussion; K. D. Brown and G. Hazlewood for cell culture facilities; T. Corps for advice on the culturing of 3T3 cells and their stimulation with growth factors; and M. J. Berridge for HPLC equipment. P.T.H. is a Lister Institute Research Fellow, and T.R.J. is Royal Society Mr. and Mrs J. Jaffé Research Fellow. L.R.S. was in receipt of a grant from Amersham International.

Human cytomegalovirus *UL97* open reading frame encodes a protein that phosphorylates the antiviral nucleoside analogue ganciclovir

Edward Littler, Amanda D. Stuart & Mark S. Chee*

Department of Molecular Sciences, Wellcome Research Laboratories, Langley Court, Beckenham, Kent BR3 3BS, UK

* Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

HUMAN cytomegalovirus (HCMV, a betaherpes virus) is the cause of serious disease in immunologically compromised individuals, including those with acquired immunodeficiency syndrome¹. One of the compounds used in the chemotherapy of HCMV infections is the nucleoside analogue 9-(1,3-dihydroxy-2-propoxymethyl)-guanine (ganciclovir). The mechanism of action of this drug is dependent on the formation of the nucleoside triphosphate, which is a strong inhibitor of the viral DNA polymerase²⁻⁴. Thymidine kinase, which is encoded by many of the herpesviruses, catalyses the initial phosphorylation of ganciclovir. But there is no evidence for the coding of this enzyme by HCMV^{2,5,6}, and DNA sequence analysis of the HCMV genome has shown that there is no open reading frame characteristic of a herpesvirus thymidine kinase⁷. Here we present biochemical and immunological evidence that the HCMV *UL97* open reading frame codes for a protein capable of phosphorylating ganciclovir. This protein seems to be responsible for the selectivity of ganciclovir and will be useful tool in the understanding and refinement of the antiviral activity of new selective anti-HCMV compounds.

In an attempt to identify the enzyme responsible for phosphorylation of ganciclovir in HCMV-infected cells, we observed that the amino-acid sequence coded by the HCMV *UL97* open reading frame (ORF) had some regions homologous to those of protein kinases. Furthermore, these sequences had homology with other proteins more distantly related to the protein kinases, such as the bacterial kinases that act on the aminoglycoside antibiotics⁸. To characterize HCMV *UL97* and investigate its ability to phosphorylate ganciclovir, we expressed the protein in a heterologous prokaryotic system. The N-terminal regions of protein kinases are non-catalytic but they do have a role in the regulation of enzymatic activity⁹⁻¹⁰. We therefore cloned a truncated part of the HCMV *UL97* ORF (positions 141, 462-142, 712 in the HCMV genome⁷) which did not code for the first 326 amino acids but retained all the sequences that align with the catalytic domains of the protein kinases. This truncated gene was cloned into the prokaryotic expression vector pGMT7 (G. Micklem, personal communication), which contains a bac-

teriophage T7 x10-s10 promoter/ribosome binding site controlling the expression of the inserted sequence¹¹. This plasmid pGMT7-*UL97tr* directs the expression of a largely insoluble protein of *M_r* 39,000 (39K), which is the predicted relative molecular mass of the open reading frame coded by the cloned DNA fragment (Fig. 1, lane 2). This protein, *UL97tr*, is not expressed in *Escherichia coli* BL21 cells containing the plasmid without the truncated *UL97*-coding insert (Fig. 1, lane 1). To confirm the viral specificity of the recombinant protein, purified *UL97tr* from BL21 was prepared by SDS-polyacrylamide gel electrophoresis and used to produce specific antisera. These sera reacted by western blotting with *UL97tr* and with a protein from HCMV-infected cells of *M_r* 80K, which is the predicted value for *UL97* (Fig. 1, lane 4). Some smaller weakly reacting proteins (lane 3) were found in infected cells; these may represent processed forms of *UL97* or degradation products.

Having demonstrated the viral specificity of *UL97tr*, we tested its ability to phosphorylate ganciclovir by assaying extracts from BL21/pGMT7-*UL97* that expressed *UL97tr* and control extracts prepared from BL21/pGMT7 but without the truncated *UL97* insert (Fig. 2). Extracts from BL21/pGMT7-*UL97* expressing *UL97tr* efficiently phosphorylated ganciclovir whereas control extracts did not. The low solubility of the *UL97tr* protein does not prevent the detection of enzymatic activity. Similar results have been found for other recombinant proteins¹² and it may be that some soluble active protein is present during the assay.

To show that the activity in these extracts was specific to recombinant *UL97tr*, we did several immunological experiments. Human sera from transplant patients having a high (post-transplant) or low (pre-transplant) titre to HCMV antigens (D. Morris, personal communication) were reacted with extracts from BL21 cells expressing *UL97tr* and the extracts assayed for residual enzyme activity (Fig. 3). The ganciclovir-phosphorylating activity in extracts of BL21/pGMT7-*UL97* expressing *UL97tr* was neutralized by the HCMV immune serum but not by the non-immune serum. Furthermore, extracts from HCMV-infected cells were assayed for their ability to phosphorylate ganciclovir either with or without prior incubation with the *UL97tr*-specific antiserum. The antiserum against *UL97tr* neutralized the kinase activity of the extracts prepared from HCMV-infected cells (Fig. 3). To verify the specificity of the kinase activity of *UL97* for ganciclovir, extracts from HCMV-infected or mock-infected cells were immunoprecipitated with antiserum against recombinant *UL97tr* and the precipitates assayed for kinase activity. The results in Fig. 4 shows that this activity immunoprecipitated with the *UL97tr* serum only from HCMV-infected cells and not from mock-infected cells. This observation is strong evidence that *UL97* has an intrinsic kinase activity towards ganciclovir.

We have shown that extracts from both HCMV-infected cells and *E. coli* expressing recombinant *UL97tr* are able to phosphorylate ganciclovir, and have provided immunological evidence that suggests that the activity in both systems is specific to the protein product of the *UL97* ORF. These results and those

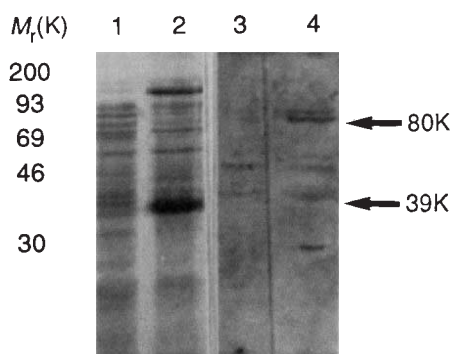


FIG. 1 Western blot analysis of *UL97tr* and of HCMV-infected cell extracts with *UL97tr*-specific antiserum. Lanes 1 and 2, proteins isolated from extracts of BL21 containing plasmid pGMT7 or plasmid expressing *UL97tr*, respectively. Extracts were prepared by sonification of BL21 cells in 50 mM glucose, 25 mM Tris-HCl, pH 8, 10 mM EDTA, 1 mM PMSF, 5 mM β -mercaptoethanol and 5 mg ml⁻¹ lysozyme, and after disruption of the cells in SDS buffer, proteins were run on 10% SDS-polyacrylamide gels. For lanes 3 and 4, proteins were prepared from MRC5 cells that had been either mock-infected or infected with HCMV strain AD169, respectively, for four days. After transfer to nitrocellulose, blots were probed using hyperimmune rabbit sera raised against purified recombinant *UL97tr* protein as described¹⁷. Molecular weight calibration is shown on the left and was derived from standard markers stained with Coomassie brilliant blue. The 39K recombinant protein and the 80K protein are indicated.

in the accompanying letter¹³, taken in conjunction with ganciclovir-resistance studies¹⁴, show that UL97 is responsible for the phosphorylation of ganciclovir in HCMV-infected cells. The finding that an HCMV-coded enzyme related in sequence to protein kinases can phosphorylate a nucleoside analogue is unexpected. UL97 is also conserved in the other herpesviruses^{8,15} and although the similarity between UL97 and the eukaryotic

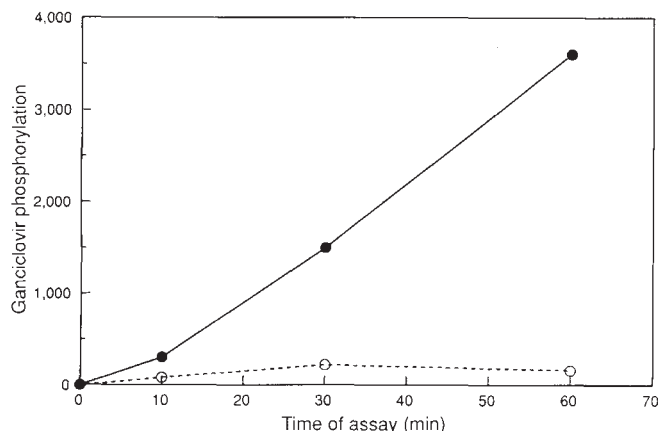


FIG. 2 Phosphorylation of ganciclovir by extracts of BL21 expressing the recombinant HCMV protein UL97tr (●), or by control extracts (○) without the UL97tr insert. Extracts of *E. coli* expressing UL97tr were prepared as described in Fig. 1 legend. Enzyme activity was assayed in 10 mM potassium phosphate, pH 8, 2 mM ATP, 2 mM dithiothreitol, 1 M NaCl, 1% nuclease-free BSA (Sigma) and 10 μ M ganciclovir (containing 15 μ Ci [3 H]ganciclovir), and the extent of phosphorylation was determined as described¹⁸. Results are expressed as pmol ganciclovir phosphorylated per min by 1 mg of bacterial extract.

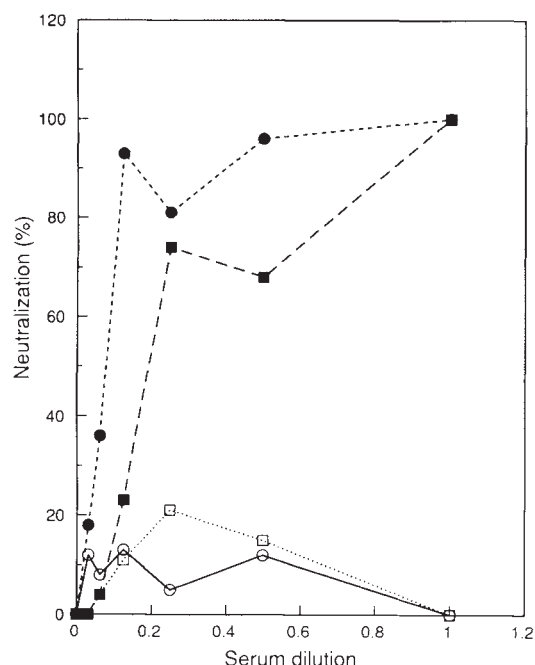


FIG. 3 Neutralization of ganciclovir kinase by antisera. Extracts prepared either from HCMV-infected MRC-5 cells or from BL21 cells containing plasmid pGMT7-UL97 were incubated with human HCMV immune (■) or non-immune serum (□), or with rabbit anti-UL97tr serum (●) or the respective preimmune control serum (○), at the dilutions indicated for 1 h at 4 °C, and then assayed for residual ganciclovir kinase activity. Results show percentage neutralization of ganciclovir kinase activity compared to enzyme to which no serum had been added.

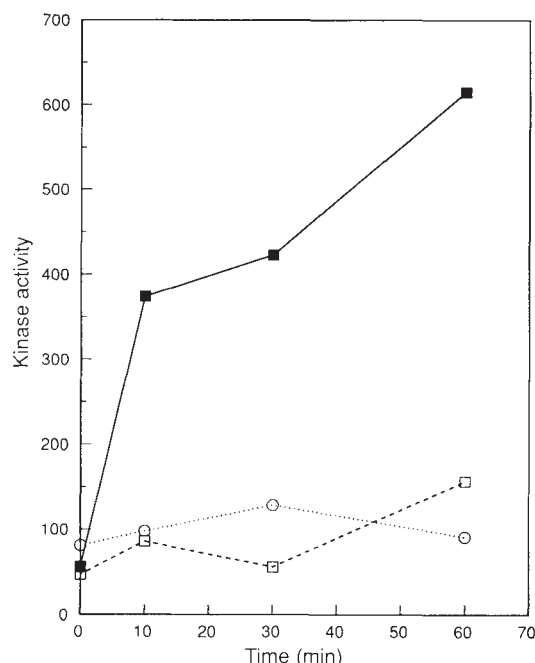


FIG. 4 Immunoprecipitation of ganciclovir kinase activity using UL97tr serum. Extracts from HCMV-infected MRC-5 cells or from mock-infected (○) cells were prepared in RIPA buffer (50 mM Tris-HCl, pH 7.5, 400 mM NaCl, 5 mM EDTA, 0.1% SDS, 0.5% NP40 and 0.5% sodium deoxycholate) and mixed with rabbit serum reacting with non-immune serum (□) or with rabbit anti-UL97tr serum (●), diluted 1:20 in PBS. The resultant immune complexes were precipitated on staph A-Sepharose beads (Pharmacia) and the pellet was washed three times with RIPA buffer. Pellets were then resuspended in enzyme assay buffer and the ganciclovir kinase activity determined.

protein kinases is low⁸, some of the other herpesvirus genes are more obviously related to the eukaryotic protein kinases. All of the herpesvirus sequences differ from the cellular enzymes at positions that are highly conserved in protein kinases. UL97 and its homologue in human herpesvirus-6 also differ significantly from the other herpesvirus sequences in regions that may be important for enzymatic activity, as determined from the crystal structure of cyclic AMP-dependent protein kinase¹⁶. One explanation for the activity seen here is that the UL97tr is modifying a bacterial protein in order to enable it to phosphorylate ganciclovir. We believe that this explanation is unlikely for the following reasons. In HCMV-infected human cells resistance to ganciclovir that is associated with reduced ganciclovir phosphorylation is due to a mutation in UL97 (ref. 13), and we have shown here that antiserum raised against UL97tr can efficiently prevent extracts from HCMV-infected human cells from phosphorylating ganciclovir. In addition, recombinant baculovirus expressing UL97 also phosphorylates ganciclovir (C. Olmstead, personal communication). These observations would demand that UL97 has the unlikely ability to modify a protein from human, insect and bacterial cells in such a way as to phosphorylate ganciclovir. It will be necessary to purify the UL97 before its ability to phosphorylate ganciclovir directly can be proven.

The role of UL97 in the replication of HCMV is not known, but unlike the herpes simplex virus homologue UL13 (D. McGeoch, personal communication), we have not yet found any protein kinase activity associated with UL97tr. One explanation may be that in the absence of a true HCMV-encoded thymidine kinase, the viral UL97 has evolved to substitute for this function. Alternatively, perhaps UL97 does function as a protein kinase, as predicted by amino-acid sequence comparison, and its ability to phosphorylate ganciclovir is a result of a chance similarity between ganciclovir and the natural substrate of the enzyme.

Our results confirm that UL97 is an unusual member of the protein kinase sequence family. Characterization of the HCMV protein responsible for phosphorylating ganciclovir will help in the design of improved antiviral nucleoside analogues for treating HCMV infection. □

Received 11 February; accepted 27 May 1992.

1. Macher, A. M. *et al.* *New Eng. J. Med.* **309**, 1454 (1983).
2. Biron, K. K. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **82**, 2473–2477 (1985).
3. Biron, K. K. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **83**, 7769–7773 (1986).
4. Freitas, V. R., Smee, D. F., Chernow, M., Boehme, R. & Matthews, T. R. *Antimicrob. Ag. Chemother.* **28**, 240–245 (1985).
5. Estes, J. E. & Huang, E. S. *J. Virol.* **24**, 13–21 (1976).
6. Fyfe, J. A., Keller, P. M., Furman, P. A., Miller, R. L. & Elion, G. B. *J. biol. Chem.* **253**, 8721–8727 (1978).
7. Chee, M. S. *et al.* *Curr. Top. Microbiol. Immun.* **154**, 125–169 (1990).
8. Chee, M. S., Lawrence, G. L. & Barrell, B. G. *J. gen. Virol.* **70**, 1151–1160 (1989).
9. Edelman, A. M., Blumenthal, D. K. & Krebs, E. G. *A. Rev. Biochem.* **56**, 567–613 (1987).
10. Pearson, R. B., Wettenhall, R. E. H., Means, A. R., Hartshorne, D. J. & Kemp, B. E. *Science* **241**, 970–973 (1988).
11. Rosenberg, A. H. *et al.* *Gene* **56**, 31–38 (1987).
12. Ertl, P. F., Thomas, M. S. & Powell, K. L. *J. gen. Virol.* **72**, 1729–1734 (1991).
13. Sullivan, V. *et al.* *Nature* **358**, 162–164 (1992).
14. Stanat, S. C. *et al.* *Antimicrob. Ag. Chemother.* **35**, 2191–2197 (1991).
15. Smith, R. F. & Smith, T. F. *J. J. Virol.* **63**, 450–455 (1989).
16. Knighton, D. R. *et al.* *Science* **253**, 414–420 (1991).
17. Littler, E., Lawrence, G., Liu, M.-Y., Barrell, B. G. & Arrand, J. R. *J. Virol.* **64**, 714–722 (1990).
18. Littler, E. & Arrand, J. R. *J. Virol.* **62**, 3892–3895 (1988).

ACKNOWLEDGEMENTS. We thank D. Morris for clinical sera and S. Stanat for help in preparing the manuscript.

A protein kinase homologue controls phosphorylation of ganciclovir in human cytomegalovirus-infected cells

V. Sullivan*, C. L. Talarico†, S. C. Stanat†, M. Davis†, D. M. Coen* & K. K. Biron†

* Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA

† Division of Virology, Burroughs Wellcome Company, Research Triangle Park, North Carolina 27709, USA

HUMAN cytomegalovirus (HCMV) is a major pathogen in immunosuppressed individuals, including patients with acquired immune deficiency syndrome. The nucleoside analogue ganciclovir (9-(1,3-dihydroxy-2-propoxymethyl)-guanine) is one of the few drugs available to treat HCMV infections, but resistant virus is a growing problem in the clinic¹ and there is a critical need for new drugs. The study of ganciclovir-resistant mutants has indicated that the selective action of ganciclovir depends largely on virus-controlled phosphorylation in HCMV-infected cells^{2–5}. The enzyme(s) responsible have not been identified. Here we report that the HCMV gene *UL97*, whose predicted product shares regions of homology with protein kinases, guanylyl cyclase and bacterial phosphotransferases^{6–8}, controls phosphorylation of ganciclovir in HCMV-infected cells. A four-amino-acid deletion of *UL97* in a conserved region, which in cyclic AMP-dependent protein kinase participates in substrate recognition⁹, causes impaired ganciclovir phosphorylation. The implications of these results for antiviral drug development and drug resistance are discussed.

To map the HCMV-encoded function that controls ganciclovir phosphorylation, the viral genome of mutant 759'D100, which is deficient in drug anabolism⁴, was cloned into an overlapping set of nine cosmids, which were then used in marker transfer experiments (Table 1). Two non-overlapping cosmids, pC7S95 and pC7S37 (Fig. 1), transferred ganciclovir resistance to the wild-type parental strain AD169. The resistance marker of pC7S37 has been mapped to the viral DNA polymerase and

TABLE 1 Marker transfer of ganciclovir resistance

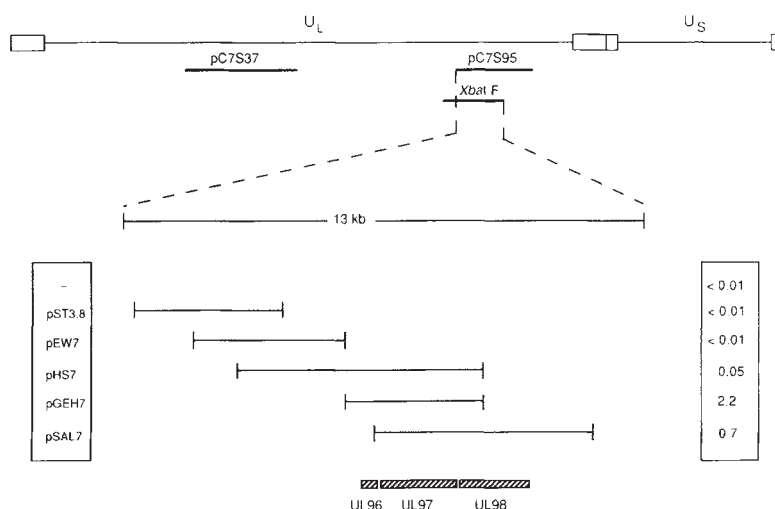
Transfected DNA	Plating efficiency (%) in ganciclovir
—	<0.04
pC7S2	<0.03
pC7S6	<0.09
pC7S11	<0.17
pC7S31	<0.04
pC7S37	2.3
pC7S40	<0.05
pC7S47	<0.23
pC7S95	2.8
pC7Sdl	<0.23

Marker transfer of ganciclovir resistance. Human foreskin fibroblasts were transfected with infectious wild-type AD169 DNA alone (—), or cotransfected with AD169 DNA and the indicated cosmid. Progeny virus from the transfected cells were tested for their ability to form plaques in 35 µM ganciclovir (per cent plating efficiency). Marker transfer of ganciclovir resistance by pC7S37 and pC7S95 is shown in bold. The preparation and maintenance of fibroblasts, preparation of infectious AD169 DNA, and details of marker transfer experiments will be described elsewhere (V.S. *et al.*, manuscript submitted). 759'D100 viral DNA was prepared essentially as described for AD169, except MRC-5 cells were used and only extracellular virus was collected. After partial digestion with *Sau3A*, fragments of 20–30 kb were ligated into the *Bam*HI site of cosmid vector pC7108 (ref. 25), packaged into phage (BRL λ packaging system), and cosmid clones isolated by phage infection of *E. coli* (strain N4956) and selected with ampicillin. Cosmids pC7S2, 6, 11, 31, 37, 40, 47, 95 and dl were isolated, characterized by digestion with restriction enzymes and Southern blot analysis (results not shown), and partially digested with *Hind*III before use in marker transfer experiments.

shown to be unrelated to ganciclovir phosphorylation (V.S. *et al.*, manuscript submitted). In contrast, the ganciclovir-resistant recombinant virus GDG'K17, isolated from marker transfer experiments with pC7S95, gave a value for the 50% effective dose of ganciclovir which was 3–10-fold higher than that of wild-type AD169 (data not shown) and was unable to induce the phosphorylation of [¹⁴C]-labelled ganciclovir in infected cells (Fig. 2a). Similarly, GDG'XBAF4, a ganciclovir-resistant recombinant virus isolated from marker transfer experiments with the F fragment of 759'D100, produced by digestion with the restriction endonuclease *Xba*I (data not shown), was also unable to phosphorylate the drug (Fig. 2a). Thus, both a ganciclovir-resistance marker of 759'D100, and the mutation responsible for the drug-phosphorylation defect are contained in the 13-kilobase (kb) overlap of the F fragment and pC7S95 (Fig. 1). Further marker transfer experiments with plasmids containing cloned 759'D100 DNA fragments within the 13-kb overlap localized the ganciclovir-resistant marker to a 2.6-kb region of DNA containing the complete *UL97* and parts of the *UL96* and *UL98* open reading frames (Fig. 1). Ganciclovir anabolism studies of the ganciclovir-resistance recombinant viruses, GDG'SAL4, GDG'HS1 and GDG'EH9, isolated from marker transfer experiments with plasmids pSAL7, pHS7 and pGEH7, respectively (Fig. 2b), confirmed that the mutation affecting ganciclovir phosphorylation was also contained in the 2.6-kb fragment. DNA sequence analysis of this region revealed only a single change when compared with the wild-type AD169 sequence: a 12-base-pair (bp) deletion within the *UL97* open reading frame (Fig. 3). These results show that the gene product of *UL97* controls ganciclovir phosphorylation in HCMV-infected cells and that mutations in this open reading frame can confer resistance to this drug.

Homologues of *UL97* are encoded by herpes simplex virus¹⁰, varicella zoster virus¹¹, Epstein-Barr virus¹² and human herpesvirus 6 (HHV-6), which suggests that there may be a conservation of function among the α, β and γ human herpesviruses⁶. These genes encode regions of homology conserved among protein kinases and phosphotransferases^{6,13,14}. The 12-bp deletion in

FIG. 1 Molecular mapping of ganciclovir-resistance markers of HCMV. The top line represents the prototype 240-kb HCMV AD169 genome¹⁸. Boxes represent inverted repeats and the solid line represents unique sequences (U_L , U_S). The solid lines below indicate the genomic position of the 759'D100 *Xba*I F fragment and cosmids pC7S37, pC7S95; the limits of the 13-kb overlap of pC7S95 and *Xba*I F are also indicated. The box on the left lists the plasmids used in the marker transfer experiments. The lines in the middle represent the 759'D100 DNA sequences within the 13-kb region contained in the indicated plasmids. The box on the right shows the plating efficiency (%) in 50 μ M ganciclovir of progeny virus from human foreskin fibroblasts transfected with AD169 DNA alone (–) or cotransfected with AD169 DNA and each of the plasmids indicated. Each of the plasmids contains DNA fragments subcloned from cosmid pC7S95. pH57 and pEW7 contain the *Hind*III S and *Eco*RI W fragments²⁶ respectively, cloned into the *Hind*III or *Eco*RI sites of pUC18. pST3.8 contains a 3.8-kb *Pst*I fragment inserted into the *Pst*I site of pGEM5Zf(+) (Promega Inc.). pGEH7 contains a 3.5-kb *Eco*RI/*Hind*III fragment inserted into the *Eco*RI and *Hind*III sites of pGEM7Zf(+). pSAL7 contains a 5.6-kb *Sal*I fragment ligated into the *Xho*I site of pGEM7Zf(+). The 759'D100 DNA fragments were excised from the plasmid vectors with the appropriate restriction enzymes before use in marker transfer experiments. Shaded box represents the 2.6-kb DNA overlap (extending from the *Sal*I to the *Hind*III sites) of



plasmids pH57, pGEH7 and pSAL7 which were able to transfer ganciclovir resistance. Hatched boxes below indicate the positions of the UL96, UL97 and UL98 open reading frames.

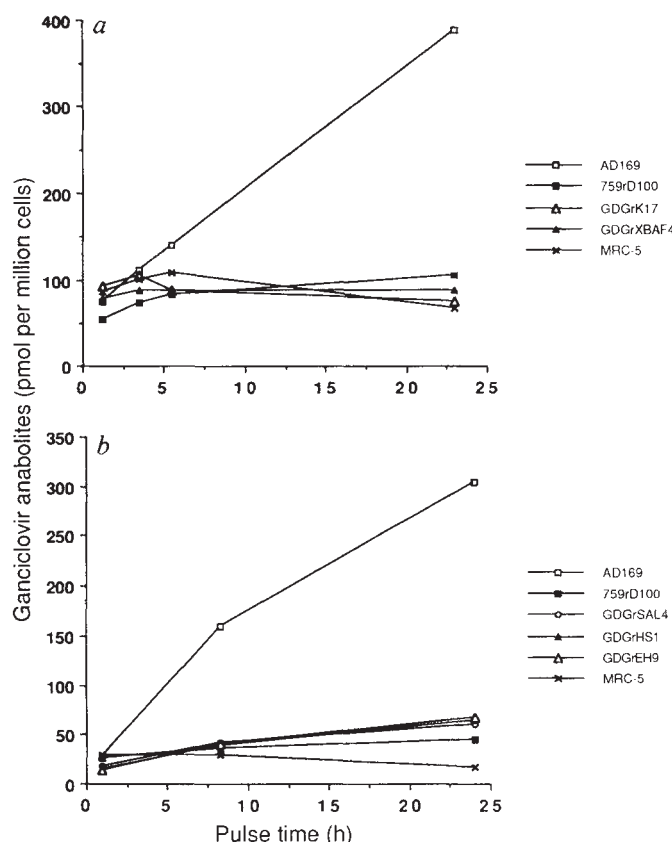


FIG. 2 Phosphorylation of 14 C-labelled ganciclovir in HCMV-infected cells. *a*, MRC-5 cells (human embryonic lung fibroblasts), maintained as described², were infected with AD169 (□), 759'D100 (■), GDG'K17 (Δ), or GDG'XBAF4 (▲) at a multiplicity of infection (m.o.i.) of 0.1–0.5, or mock-infected (×). Four days post-infection they were pulse-labelled for 1.25, 3.5, 5.5 and 23 h with 50 μ M 14 C-labelled ganciclovir (specific activity, 52mCi per mmol) and, after extraction with perchloric acid, the ganciclovir anabolites were measured with a cation-exchange column as described²⁷. *b*, MRC-5 cells were infected with AD169 (□), 759'D100 (■), or GDG'SAL4 (○), GDG'HS1 (▲), GDG'EH9 (△), at m.o.i. 0.1–0.5, or mock infected (×). Four days post-infection they were labelled with 14 C-labelled ganciclovir for 1, 8.3 and 24 h, and the levels of ganciclovir anabolites determined.

UL97 results in a 4-amino-acid deletion in the UL97 product of residues 638–641 (Ala-Ala-Cys-Arg) (Fig. 3). These residues lie in a region conserved among protein kinases, designated subdomain IX (ref. 14) or region VI (ref. 13). The conservation of this region indicates that it may have an important and common function in these enzymes^{13,14}. X-ray crystallography of the catalytic subunit of bovine cAMP-dependent protein kinase has shown that residues corresponding to those deleted in UL97 form a hydrophobic binding site for a peptide inhibitor (PKI (5–24)) and, by extension, its substrate⁹. The deletion of residues at positions implicated in substrate binding of protein kinases raises the possibility that the mutant 759'D100 UL97 product may retain enzyme activity with altered substrate specificity. It is interesting to note that three of the deleted amino acids are conserved in the HHV-6 homologue of UL97 (15R in Fig. 3). HHV-6 is also sensitive to ganciclovir^{15,16} and, like HCMV, no homologue of the thymidine kinase responsible for ganciclovir phosphorylation in cells infected by herpes simplex virus¹⁷ has been found in this virus^{15,16,18}. This raises the possibility that the HHV-6 homologue of UL97 may be responsible for ganciclovir phosphorylation in cells infected by HHV-6.

The role of UL97 in the HCMV replication cycle is not known. The predicted product is obviously related to protein kinases and kinase activity has been reported in HCMV-infected cells and virions^{19–23}, but no protein kinase function has yet been associated with UL97.

On the basis of these results, we envisage two possible mechanisms by which ganciclovir phosphorylation is controlled by UL97. UL97 may phosphorylate another enzyme, thereby activating this second enzyme to phosphorylate ganciclovir; although we cannot exclude this model, we favour a different mechanism in which UL97 itself phosphorylates the drug. This simpler mechanism is plausible because UL97 departs from consensus protein kinase motifs at the same residues that also vary in bacterial aminoglycoside phosphotransferases⁷ and in sea urchin integral membrane guanylyl cyclase⁸, enzymes that transfer phosphates to sugar hydroxyl moieties. Ganciclovir is phosphorylated on its acyclic sugar-like moiety. Indeed, Chee *et al.*, who pointed out these divergences, have suggested that UL97 deserves consideration as the HCMV-encoded ganciclovir kinase⁶, a proposal supported by evidence in the accompanying paper²⁴ showing that UL97 can carry out this enzymatic function.

The mapping of an HCMV ganciclovir-resistance marker to UL97 identifies a new selective target which should facilitate the development of safer and more effective anti-HCMV drugs.

Resistance of HCMV to ganciclovir in the clinic is predominantly due to impaired drug phosphorylation⁵, presumably as a result of UL97 mutation. It is important to determine the role of UL97 in the pathogenicity of the virus. If UL97 proves to be essential to the virus, then ganciclovir or a new drug might act not only as a substrate for this enzyme but also as an inhibitor of its normal function in the virus replication cycle. Independently, UL97 may help define residues and structures important for catalytic specificity during the evolution of a protein kinase homologue into an enzyme with a different function. □

1. Drew, W. L. *et al. J. infect. Dis.* **163**, 716-719 (1991).
2. Biron, K. K. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 2473-2477 (1985).
3. Freitas, V. R., Smee, D. F., Chernow, M., Boehme, R. & Matthews, T. R. *Antimicrob. Ag. Chemother.* **28**, 240-245 (1985).
4. Biron, K. K. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 8769-8773 (1986).
5. Stanat, S. C. *et al. Antimicrob. Ag. Chemother.* **35**, 2191-2197 (1991).
6. Chee, M. S., Lawrence, G. L. & Barrell, B. G. *J. gen. Virol.* **70**, 1151-1160 (1989).
7. Brenner, S. *Nature* **329**, 21 (1987).
8. Singh, S. *et al. Nature* **334**, 708-712 (1988).
9. Knighton, D. R. *et al. Science* **253**, 414-420 (1991).
10. McGeoch, D. J. *et al. J. gen. Virol.* **69**, 1531-1574 (1988).
11. Davison, A. J. & Scott, J. E. *J. gen. Virol.* **67**, 1759-1816 (1986).
12. Baer, R. *et al. Nature* **310**, 207-211 (1984).
13. Smith, R. F. & Smith, T. F. *J. Virol.* **63**, 450-455 (1989).
14. Harks, S. K., Quinn, A. M. & Hunter, T. *Science* **241**, 42-52 (1988).
15. Agut, H. *et al. Res. Virol.* **140**, 219-228 (1989).
16. Di Luca, D., Katsafanas, G., Schirmer, E. C., Balachandran, N. & Frenkel, N. *Virology* **175**, 199-210 (1990).
17. Fyfe, J. A., Keller, P. M., Furman, P. A., Miller, R. L. & Elion, G. B. *J. biol. Chem.* **253**, 8721-8727 (1978).
18. Chee, M. S. *et al. Curr. Top. Microbiol. Immun.* **154**, 125-169 (1990).
19. Mar, E.-C., Patel, P. C. & Huang, E.-S. *J. gen. Virol.* **57**, 149-156 (1981).
20. Michelson, S., Tardy-Panit, M. & Barzo, O. *Virology* **134**, 259-268 (1984).
21. Roby, C. & Gibson, W. *J. Virol.* **59**, 714-727 (1986).
22. Britt, W. J. & Auger, D. *J. Virol.* **59**, 185-188 (1986).
23. Somogyi, T., Michelson, S. & Masse, M.-J., O. *Virology* **174**, 276-285 (1990).
24. Littler, E., Stuart, A. D. & Chee, M. S. *Nature* **358**, 160-162 (1992).
25. Lau, Y.-F. & Kan, Y. W. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5225-5229 (1983).
26. Fleckenstein, B., Muller, I. & Collins, J. *Gene* **118**, 39-46 (1982).
27. Fyfe, J. A., McKee, S. A. & Keller, P. M. *Molec. Pharmacol.* **24**, 316-323 (1983).
28. Shoji, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 848-851 (1981).

Structure of astacin and implications for activation of astacins and zinc-ligation of collagenases

The five-stranded β -sheet and both helices of the 'N' domain of astacin are topologically similar to thermolysin, although the

TABLE 1 Sequence alignments

Astacin numbering			92	93				96				99				102	103					
Astacin	T	I	I	H	E	L	M	H	A	I	G	F	Y	H	E	H	T	R	M	D	R	D
BMP1	I	V	V	H	E	L	G	H	V	V	G	F	W	H	E	H	T	R	P	D	R	D
TOL	I	I	T	H	E	L	G	H	T	I	G	F	H	H	E	H	A	R	G	D	R	D
SPAN/BP10	T	I	V	H	E	I	G	H	A	I	G	F	H	H	E	Q	S	R	P	D	R	D
Meprin A	T	I	E	H	E	I	L	H	A	L	G	F	F	H	E	Q	S	R	T	D	R	D
PPH	I	I	E	H	E	I	L	H	A	L	G	F	Y	H	E	Q	S	R	T	D	R	D
Stromelysin 1	V	A	A	H	E	I	G	H	S	L	G	L	F	H	S	A	N	T	E	A	L	M
Haemorrhagictoxin-d	T	M	A	H	E	L	G	H	N	L	G	M	E	H	D	G	K	D	C	L	R	G
Serratia <i>Proteinase</i>	T	F	T	H	E	I	G	H	A	L	G	L	S	H	P	G	D	Y	N	Y	G	E
Thermolysin	V	V	G	H	E	L	T	H	A	V	T	D	Y	T	A	G	L	V	Y	Q	N	E

Sequence alignment¹¹ (shown in one-letter-code) around the zinc-binding consensus sequence for astacin² and the related metalloproteinases of the astacin family, namely human bone morphogenetic protein (BMP1)⁶, the dorsal-ventral patterning gene-product tolloid from fruitfly (TOL)⁷, the embryonal protein from sea urchin (SPAN and BP10)^{9,10}, meprin A (from renal mice brush-border membrane⁵), and human PABA-peptide hydrolase from intestinal brush-border (PPH)⁵. The overall sequence similarity of the 200-amino-acid molecule astacin with the catalytic domains of these homologous family members is about 35%. In addition, one representative proteinase each of the vertebrate collagenases (human stromelysin-1²³), of the snake venom zinc-proteinases (haemorrhagic toxin-d from *Crotalus atrox*²⁴), and of the large bacterial proteinases (*Serratia* proteinase²⁵), which likewise contain the consensus sequence HEXHXXGXXH, is displayed. The shorter consensus sequence of thermolysin²⁶ is also shown. The incomplete cDNA of the embryonic developmental protein UVS.2 from frog⁸ comprises only a homologous segment to the C domain in astacin and is therefore not included in the alignment.

FIG. 1 Astacin structure *a*, displayed together with a Connolly dot surface (with a rolling sphere of 1.4 Å radius, using the program MAIN²⁵), or *b*, shown as a ribbon model (using the program RIBBON²⁶). An upper (N) and a lower (C) domain are separated by a deep active-site cleft, accommodating the zinc ion (displayed as a green-dotted sphere (*a*) or a red sphere (*b*) of radius 1.0 Å). The protein was crystallized in trigonal crystals of space group *P*3₁21 (cell constants *a* = *b* = 61.96 Å, *c* = 98.52 Å) from ammonium sulphate pH 7.0. X-ray intensity data from a native crystal (to 1.8 Å) and from six heavy-atom derivatives (to 2.2 Å) were collected on a FAST television area detector diffractometer (Enraf Nonius, Delft), evaluated on-line by means of MADNES²⁷ and processed. Heavy-atom positions were determined and refined yielding an overall figure-of-merit of 0.71 (20 to 2.2 Å). An astacin starting model built against an initial 3 Å 2*F*_o - *F*_c multiple isomorphous replacement map was improved and completed in iterative cycles, each comprising energy-restrained crystallographic refinement with X-PLOR²⁸ and modelling (FRODO²⁹) against Fourier maps calculated with combined or (finally) model calculated phases. The crystallographic *R*-value of the final model consisting of 1,593 non-hydrogen protein atoms and 181 localized solvent molecules is 0.162 for 19,805 independent reflections between 10 and 1.8 Å; the r.m.s. bond derivation from target values is 0.012 Å.

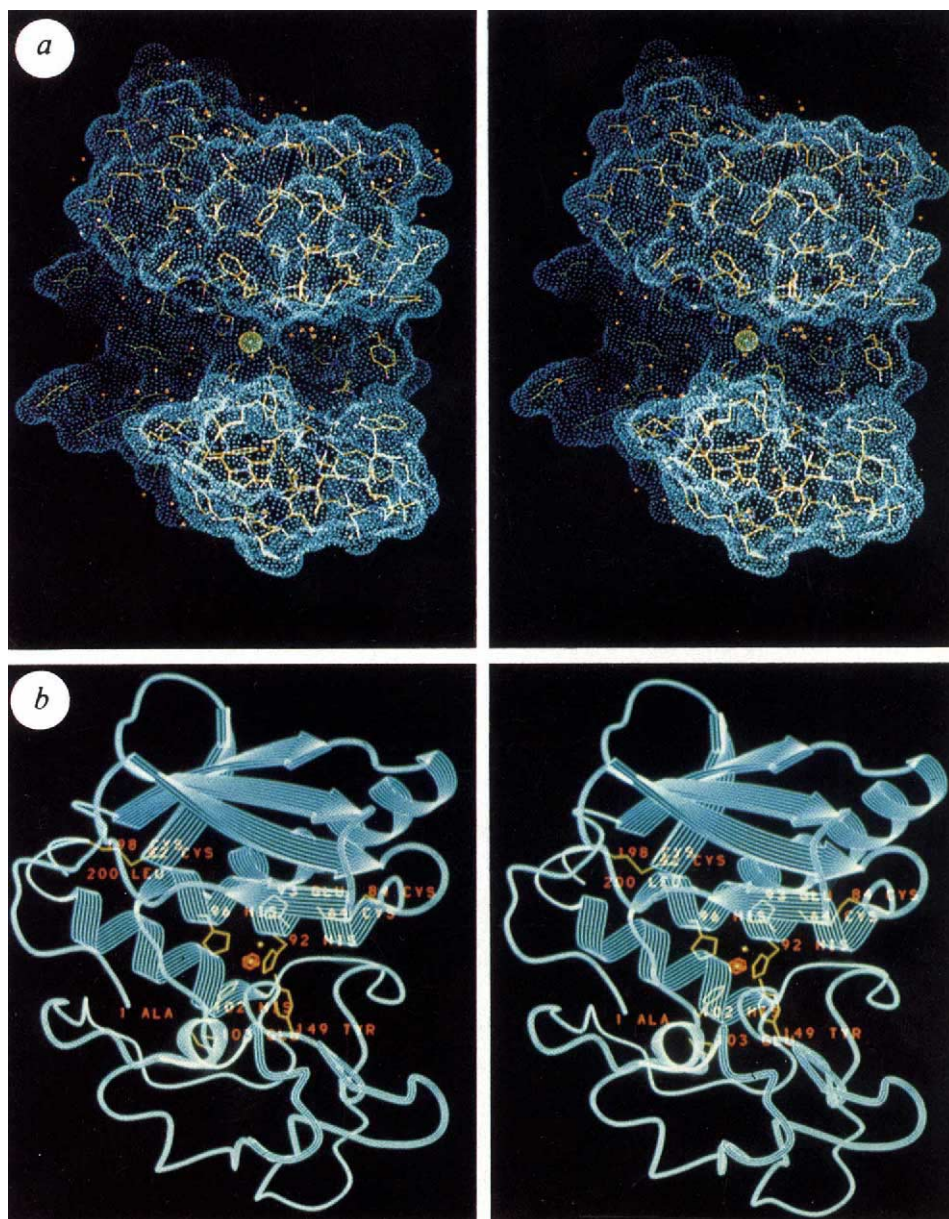
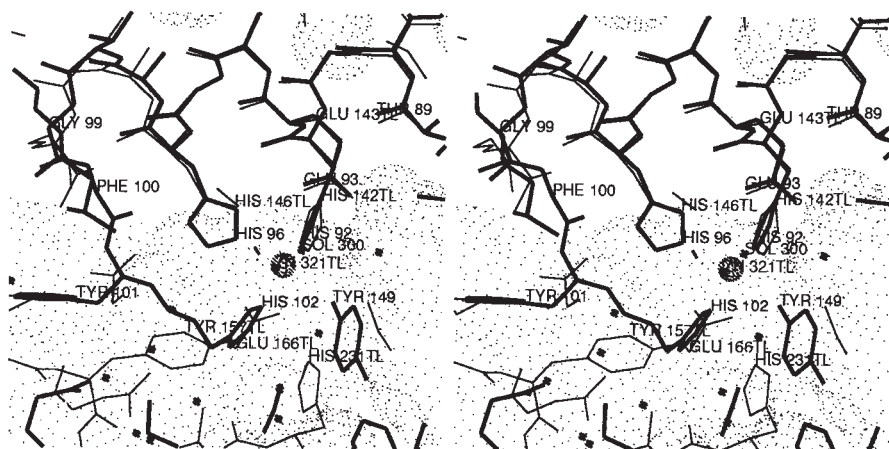


FIG. 2 Comparison of the zinc-binding sites of astacin and thermolysin (using MAIN²⁵). Orientation as in Fig. 1. The astacin structure (that is, the complete main-chain from Thr 89 to His 102 and the five zinc-liganding side-chains of His 92, Glu 93 (including SOL 300), His 96, His 102 and Tyr 149, bold connections) is displayed together with its Connolly dot surface, the zinc ion (dense sphere of radius 0.5 Å) and a few localized solvent molecules (crosses). At Gly 99, the astacin chain exits from the 'active-site helix' and turns 'downwards' (centre). It remains on the surface until His 102, but is buried from Glu 103 onwards. The structure of non-ligated thermolysin¹² (Protein Data Bank code 3TLN³⁰, thin connections) is superimposed after a least-squares fit of its central active-site helix residues 137TL–146TL to astacin segment His 87–His 96; the zinc-ion of thermolysin is labelled with ZN321TL.



two proteins have only 5 out of 200 sequence positions in common (Table 1). In contrast, the N-terminal strap and the organization of the lower domain are completely unrelated. Astacin segment Cys 64–Val 68 forms the upper edge of the active-site cleft in a similar fashion to the corresponding edge β -strand in thermolysin, so that the carbonyl group of Cys 64 presumably plays a similar role in hydrogen-bonding and polarizing the amino group of the P1' residue of a bound substrate¹⁶. By further analogy to thermolysin, Glu 93 and the zinc ion probably polarize water molecule Sol 300, promoting its attack on the scissile peptide bond. Tyr 149, occupying a similar site to His 231TL in thermolysin, might likewise assist by hydrogen-bonding to the P1–P1' carbonyl group. The central zinc, the active-site helix Tyr 86–Gly 99 and the zinc-liganding residues His 92, Glu 93 and His 96 of astacin may be superimposed with respective features of thermolysin. But from Gly 99 onwards astacin exhibits a unique conformation lacking any similarity to other proteinases of known tertiary structure, and His 102 of astacin is replaced by a glutamic acid (Glu 166) in thermolysin, whose zinc is tetrahedrally coordinated.

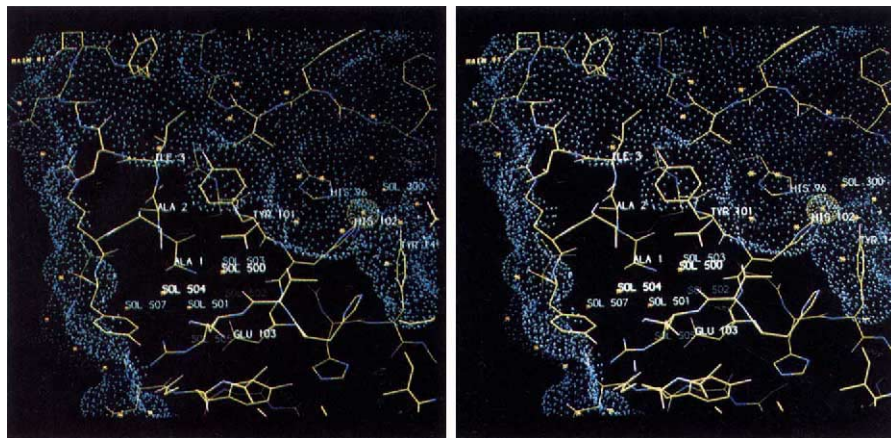
The positioning of the three histidine residues and the conservation of most of the intervening residues in the consensus sequence (Table 1) indicate a similar spatial arrangement of these residues around the active-site zinc ion not only for the members of the astacin family, but also for the vertebrate collagenases, snake venom proteinases and large bacterial proteinases. In particular astacin residue Gly 99, which exhibits a main-chain conformation available to glycine residues only, explains its absolute conservation. Moreover, the identification of the 'third imidazole zinc-ligand' in these various metalloendopeptidases, corresponding to His 102 of astacin, indicates a solution to a long-standing problem¹⁷. But lack of sequence

conservation C-terminal to the third histidine residue suggests considerable differences in the active-site cleft of these non-astacin-like proteinases.

A glutamic acid (Glu 103), following the 'third' histidine (His 102), is found only in the astacin family members (Table 1). This suggests that the burial and interactions of the glutamic acid carboxylate group (Glu 103) with the N-terminal ammonium group (Fig. 3) are similar for all mature astacin-like proteinases. An N-terminally elongated chain, such as for a pro-form, could not be accommodated in such a manner; the N terminus would rather have to be placed outside the proteolytic domain, thereby destabilizing the structure and possibly affecting the catalytic site and substrate-binding region (Figs 2 and 3).

The astacin-homologues meprin A and B, presumably isolated as mature and 'latent' zinc-endopeptidases, respectively, can be isolated from mouse renal brush-border membranes¹⁸. Limited trypsinolysis of meprin B (which is less thermostable than meprin A) yields a 5–20-fold more active species, indicative of a specific activation mechanism. No pro-form has yet been found for astacin; partial complementary DNA information suggests, however, an upstream peptide elongation of the virgin protein chain (L. T. Jacob, R.Z. and W.S., unpublished results). According to peptide and cDNA sequence information, N-terminally elongated forms of all members of the astacin family (except UVS.2; ref. 8 and Table 1) exist; the N termini of astacin and those of the mature form of PABA-peptide hydrolase and the papain-purified meprin A start at equivalent sequence positions (with an alanine or an asparagine, respectively⁵).

The burial of the N terminus in astacin and the presumed conformational changes of proteolytic activation of pro-forms are reminiscent of the trypsin-related serine proteinases¹⁹, where



the N-terminal segment of the active enzyme inserts into the molecular body, with its ammonium group forming a buried salt-bridge with an acidic residue adjacent to an active-site residue. Activation cleavage liberates the mature enzyme N terminus, triggering the conformational transition from a looser pro-form to the more compact active state. The presumed similar 'astacin'-activation mechanism, by contrast, very much differs from the 'cysteine switch'²⁰ or 'velcro'¹⁷ activation mechanism suggested for the vertebrate collagenases where the pro segment presumably folds across the active-site, with a conserved cysteine coordinating the zinc. □

Received 16 March; accepted 22 May 1992.

1. Stöcker, W., Wolz, R. L., Zwilling, R., Strydom, D. J. & Auld, D. S. *Biochemistry* **27**, 5026–5032 (1988).
2. Titani, K. et al. *Biochemistry* **26**, 222–226 (1987).
3. Stöcker, W., Sauer, B. & Zwilling, R. *Biol. Chem. Hoppe-Seyler* **372**, 385–392 (1991).
4. Barrett, A. J. & Rawlings, D. *Biochem. Soc. Trans.* **19**, 707–715 (1991).
5. Dumermuth, E. et al. *J. Biol. Chem.* **266**, 21381–21385 (1991).
6. Wozney, J. M. et al. *Science* **242**, 1528–1534 (1988).
7. Shimell, M. J., Ferguson, E. L., Childs, S. R. & O'Connor, M. B. *Cell* **67**, 469–481 (1991).
8. Sato, S. M. & Sargent, T. D. *Dev. Biol.* **137**, 135–141 (1990).
9. Reynolds, S. D., Angerer, L. M., Palis, J., Nasir, A. & Angerer, R. C. *Development* **114**, 769–786 (1992).

10. Lepage, T., Ghiglieri, C. & Cache, C. *Development* **114**, 147–164 (1992).
11. Stöcker, W., Ng, M. & Auld, D. S. *Biochemistry* **29**, 10418–10425 (1990).
12. Matthews, B. W., Jansonius, J. N., Colman, P. M., Schoenborn, B. P. & Dupourque, D. *Nature new Biol.* **238**, 37–41 (1972).
13. Paupit, R. A. et al. *J. molec. Biol.* **199**, 525–537 (1988).
14. Thayer, M. M., Flaherty, K. M. & McKay, D. B. *J. Biol. Chem.* **266**, 2864–2871 (1991).
15. Lipscomb, W. N. et al. *Brookhaven Symp. Biol.* **21**, 24–90 (1968).
16. Matthews, B. W. *Accts chem. Res.* **21**, 333–340 (1988).
17. Vallee, B. L. & Auld, D. S. *Biochemistry* **29**, 5647–5659 (1990).
18. Kounnas, M. Z., Wolz, R. L., Gorbea, C. M. & Bond, J. S. *J. Biol. Chem.* **266**, 17530–17537 (1991).
19. Huber, R. & Bode, W. *Accts chem. Res.* **11**, 114–122 (1978).
20. Springman, E. B., Angleton, E. L., Birkeid-Hansen, H. & Van Wart, H. E. *Proc. natn. Acad. Sci. U.S.A.* **87**, 364–368 (1990).
21. Goldberg, G. J. et al. *J. Biol. Chem.* **261**, 6600–6605 (1986).
22. Shannon, J. D., Baramova, E. N., Bjarnason, J. B. & Fox, J. W. *J. Biol. Chem.* **264**, 11575–11583 (1989).
23. Whitham, S. E. et al. *Biochem. J.* **240**, 913–916 (1986).
24. Titani, K., Hermodsson, M. A., Ericsson, L. H., Walsh, K. A. & Neurath, H. *Nature new Biol.* **238**, 35–37 (1972).
25. Turk, D. thesis, Technische Universität, München (1992).
26. Priestley, J. P. *J. appl. Crystallogr.* **21**, 572–576 (1988).
27. Messerschmidt, A. & Pflugrath, J. W. *J. appl. Crystallogr.* **20**, 306–315 (1987).
28. Brünger, A. T. *X-PLOR Manual version 2.1* (Yale University, New Haven, 1990).
29. Jones, T. A. *Acta Crystallogr. A* **34**, 931–935 (1978).
30. Bernstein, F. C. et al. *J. molec. Biol.* **122**, 535–542 (1977).

ACKNOWLEDGEMENTS. We thank W. Forster for excellent help in protein purification, I. Mayr in crystallization, and M. Stubbs for reading this manuscript. The financial support of the SFB207 of the Universität München, of the Fond der Chemischen Industrie and of the Deutsche Forschungsgemeinschaft are acknowledged.

Shigella flexneri induces apoptosis in infected macrophages

Arturo Zychlinsky, Marie Christine Prevost* & Philippe J. Sansonetti†

Unité de Pathogénie Microbienne Moléculaire, INSERM U199 and
* Station Centrale de Microscopie Electronique, Institut Pasteur,
25–28 rue du Docteur Roux, 75724 Paris Cedex 15, France

THE Gram-negative bacterial pathogen *Shigella flexneri* causes dysentery by invading the human colonic mucosa¹. Bacteria are phagocytosed by enterocytes², escape from the phagosome into the cytoplasm³ and spread to adjacent cells⁴. After crossing the epithelium, *Shigella* reaches the lamina propria of intestinal villi, the first line of defence. This tissue is densely populated with phagocytes that are killed in great numbers, resulting in abscesses^{5,6}. The genes required for cell invasion⁷ and macrophage killing² are located on a 220-kilobase plasmid. We report here on the mechanism of cytotoxicity used by *S. flexneri* to kill macrophages. Each of four different strains was tested for its capacity to induce cell death. An invasive strain induced programmed cell death (apoptosis^{8,9}), whereas its non-invasive, plasmid-cured isogenic strain was not toxic; neither was a mutant in *ipa B* (ref. 10) (invasion protein antigen), a gene necessary for entry. A non-invasive strain expressing the haemolysin operon of *Escherichia coli*¹¹ induced accidental cell death (necrosis), demonstrating that other bacterial cytotoxic mechanisms do not lead to apoptosis. This is the first evidence that an invasive bacterial pathogen can induce suicide in its host cells.

S. flexneri cytotoxicity was tested on the murine macrophage cell line J774 in a standard ⁵¹Cr-release assay¹². We tested wild-type strain M90T, a *S. flexneri* serotype 5 isolate carrying the 220-kilobase (kb) plasmid pWR100 that encodes the invasive phenotype⁷, and three isogenic derivatives of M90T: SC403, an *ipa B* insertion mutant, non-invasive strain¹⁰; BS176, a plasmid-less, non-invasive strain; and BS176-hyl, which is BS176 transformed with pANN202-312, a recombinant plasmid expressing the plasmid-borne *E. coli* α -haemolysin cistrons¹¹. M90T killed macrophages effectively, with maximal cytotoxicity after 3 h incubation. In contrast, the non-invasive SC403 and BS176 were not at all cytotoxic. But the non-invasive, haemolysin-expressing strain BS176-hyl destroyed J774 macrophages consistently faster

than M90T-induced killing, implying a more immediate mechanism of toxicity (Fig. 1a).

To determine whether the bacteria needed to be within the cytoplasm to induce cell death, we prevented phagocytosis by pretreating J774 macrophages with cytochalasin D, an inhibitor of actin polymerization¹³. Under these conditions, cell death caused by M90T, but not by BS176-hyl, was almost completely

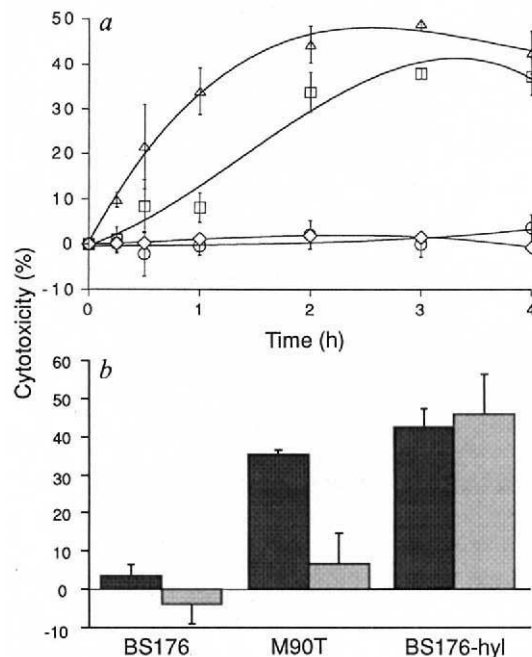


FIG. 1 Cytotoxicity of *S. flexneri* strains on the J774 macrophage cell line. a, Time course of ⁵¹Cr-release assay of the invasive strain M90T (□), a non-invasive *ipa B* insertion mutant SC403 (◇), a non-invasive derivative BS176 (○), and BS176 transfected with the *E. coli* haemolysin operon BS176-hyl (Δ). b, ⁵¹Cr-release at 4 h after infection of the indicated strains on J774 cells (dark boxes) and J774 cells preincubated with cytochalasin D (light boxes).

METHODS. J774 cells were labelled with Na₂⁵¹CrO₄ (Amersham), washed and seeded onto 96-well plates¹². After the cells had attached, they were infected with the indicated strain at a multiplicity of infection of 150 bacteria per cell as described elsewhere². At the appropriate time points the ⁵¹Cr present in the supernatant was measured. Cytochalasin D was added 30 min before infection at a concentration of 1 μg ml⁻¹. Error bars represent the standard deviations of experiments done in triplicate.

† To whom correspondence should be addressed.

FIG. 2 Electron micrographs of J774 cells infected with different *S. flexneri* strains. *a, b*, Ninety minutes after infection with M90T, showing apoptotic morphology: intense perinuclear chromatin aggregation, cytoplasmic blebbing, cytoplasmic vacuolization and maintenance of organelle structure. *c*, Incubation in serum-free medium for 6 h with apoptotic morphology similar to M90T-infected cells. *d*, Thirty min after infection with BS176-hyl, showing necrotic morphology: flocculation of the chromatin, absence of recognizable organelles and dissolution of the cytoplasm. *e*, Ninety min after infection with BS176, showing healthy cells containing bacteria within vacuoles. *f*, Incubation for 6 h in serum-containing medium, as a control. J774 cells were infected at a multiplicity of infection of 150 bacteria per cell and, at the times indicated, fixed *in situ*, embedded, thin-sectioned and observed by electron microscopy, as previously described³. Scale bars, 1 μ m.

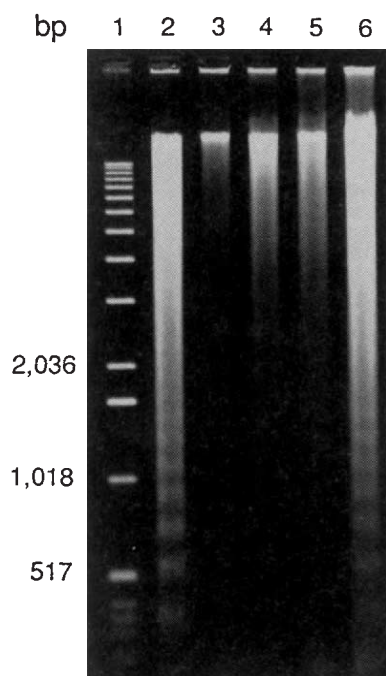
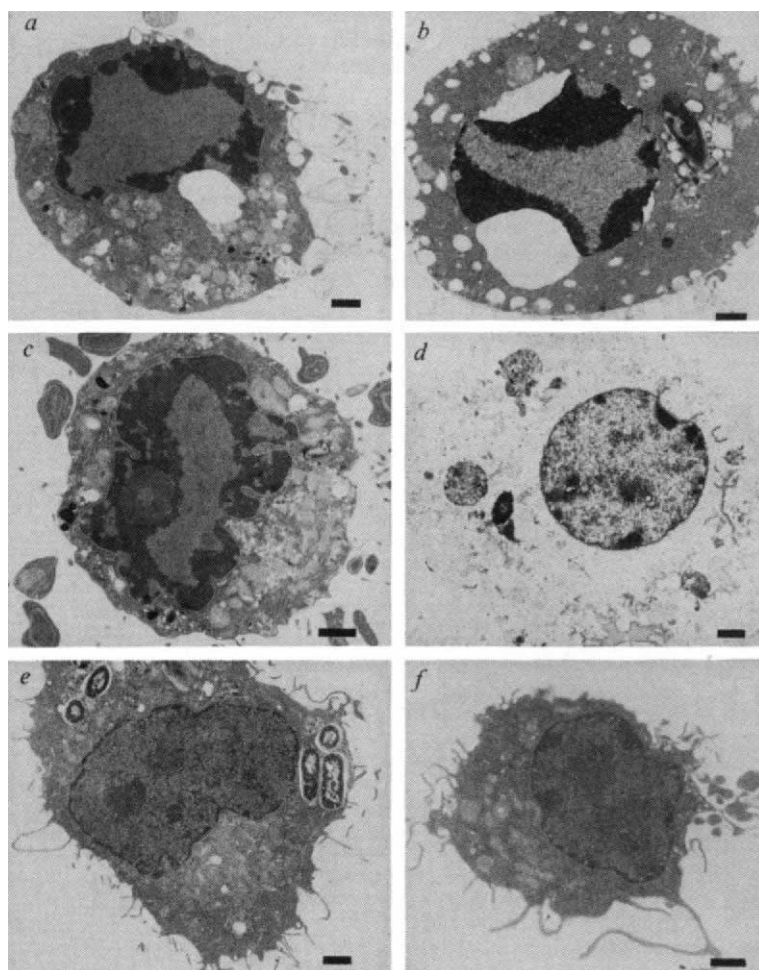


FIG. 3 DNA fragmentation of J774 cells. Cell lysis and DNA electrophoresis were done as described¹². Lanes 1, 1-kb ladder molecular size markers (BRL); 2, serum-starved cells after 6 h; 3, cells incubated in the presence of serum for 6 h; 4, cells infected with BS176 for 3 h; 5, cells infected with BS176-hyl for 1 h; 6, cells infected with M90T for 3 h. Only serum-starved and M90T-infected cells generated ladders of 200 bp characteristic of apoptosis. Cells infected with BS176-hyl were negative for DNA fragmentation at shorter and longer incubation periods.

METHODS. After 1 h of infection with 7×10^5 cells (ref. 2) or serum-starvation, cells were washed and further incubated in the presence of $50 \mu\text{g ml}^{-1}$ gentamycin. At the indicated times, cells were lysed¹² in 8 mM EDTA, 10 mM Tris, pH 7.2, 0.2% Triton. The postnuclear supernatant was extracted with phenol and phenol-chloroform before precipitation with ethanol. The DNA solution was incubated for 2 h in the presence of $1 \mu\text{g ml}^{-1}$ RNase and electrophoresed in a 1% agarose gel. The gel was then stained with ethidium bromide.

abolished (Fig. 1b). These data demonstrate that M90T is cytotoxic only when it is inside the cytoplasm, whereas BS176-hyl can destroy cells from the outside. Thus it is likely that the *E. coli* haemolysin, which is a pore-forming protein, kills cells by disrupting their cytoplasmic membrane¹⁴.

Apoptotic and necrotic cells can be distinguished by their different ultrastructural morphologies⁸. Using electron microscopy, we demonstrated that cells infected with M90T showed clear manifestations of apoptosis: condensation of chromatin at the nuclear boundary; blebbing at the cell surface; cytoplasmic vacuolization; dilation of the endoplasmic reticulum; and preservation of organelle structure (Fig. 2a, b). A similar morphology is observed in J774 cells incubated in the absence of serum for more than 6 h, a known signal for apoptosis¹⁵ (Fig. 2c). In contrast, as early as 30 min after infection, cells incubated with BS176-hyl showed clear signs of necrosis: disappearance of recognizable organelles from the cytoplasm, flocculation of the chromatin, but conservation of the shape of nucleus (Fig. 2d). J774 cells infected with BS176 had a normal, healthy appearance, and contained bacteria within vacuoles (Fig. 2e, f).

A universal characteristic of apoptosis seems to be the cleavage of the DNA of the dying cell at the internucleosomal regions, resulting in DNA fragments which are multimers of 200 base pairs (bp). DNA fragmentation is mediated by a cellular endonuclease^{8,9}. Infection with M90T induced DNA fragmentation in J774 macrophages, whereas infection with BS176 or BS176-hyl did not (Fig. 3). Similar results were obtained with murine peritoneal macrophages and with the human monocyte cell line U937 differentiated for 48 h with phorbol myristate acetate (data not shown). As a positive control, DNA from serum-deprived, apoptotic cells¹⁵ presented an identical pattern to that of cells infected with M90T (Fig. 3). Infection of J774

cells with *Listeria monocytogenes*, a Gram-positive pathogen that also invades the cytoplasm^{16,17}, did not cause DNA fragmentation (data not shown), thereby demonstrating that bacterial invasion of the cytoplasm does not necessarily lead to apoptosis. Incubation of J774 cells with the protein synthesis inhibitor cycloheximide or emetine in conditions that abrogated *de novo* ³⁵S-methionine incorporation by more than 95% did not suppress cell death induced by M90T as measured using ⁵¹Cr-release assays (data not shown). This latter result agrees with previous observations that apoptosis in macrophages is independent of macromolecular synthesis¹⁸.

The massive influx and subsequent death of inflammatory cells in the colonic submucosa is a prominent feature of *Shigella* infections^{5,6}. Our data are consistent with an essential role of apoptosis in shigellosis. The identification of a bacterial system that elicits eukaryotic programmed cell death represents a novel pathogenic mechanism of intracellular bacteria. Future experiments to determine whether *S. flexneri* activates the intrinsic suicide program in the host cell or short-circuits the program by producing a factor that directly causes cell death should provide general insights into the molecular mechanisms of apoptosis. □

Received 28 February; accepted 27 May 1992.

1. LaBrec, E. H., Schneider, H., Magnani, T. J. & Formal, S. B. *J. Bact.* **88**, 1503-1518 (1964).
2. Clerc, P. L., Ryter, A., Mounier, J. & Sansonetti, P. J. *Infect. Immun.* **55**, 521-527 (1987).
3. Sansonetti, P. J., Ryter, A., Clerc, P., Maurelli, A. T. & Mounier, J. *Infect. Immun.* **51**, 461-469 (1986).
4. Bernardini, M. L., Mounier, J., d'Hauteville, H., Coquis-Rondon, M. & Sansonetti, P. J. *Proc. natn. Acad. Sci. U.S.A.* **86**, 3867-3871 (1989).
5. Anand, B. S. *et al. Gastroenterology* **90**, 654-660 (1986).
6. Mathan, M. M. & Mathan, V. I. *Rev. infect. Dis.* **13** (suppl. 4), 314-318 (1991).
7. Sansonetti, P. J., Kopecko, D. J. & Formal, S. B. *Infect. Immun.* **35**, 852-860 (1982).
8. Arends, M. J. & Wyllie, A. H. *Int. Rev. exp. Path.* **32**, 223-254 (1991).
9. Ellis, R. E., Yuan, J. & Horvitz, H. R. *Rev. Cell Biol.* **7**, 663-698 (1991).
10. High, N., Mounier, J., Prevost, M. C. & Sansonetti, P. J. *EMBO J.* **11**, 1991-1999 (1992).
11. Noegel, A., Rdest, U., Springer, W. & Goebel, W. *Molec. Gen. Genet.* **175**, 343-350 (1979).
12. Zheng, L. M., Zychlinsky, A., Liu, C.-C., Ojcius, D. M. & Ding-F. Young, J. *J. Cell Biol.* **112**, 279-288 (1991).
13. Tanenbaum, S. W. *Cytochalasin: Biochemical and Cell Biological Aspects* (North-Holland, Amsterdam, 1978).
14. Welch, R. A. *Molec. Microbiol.* **5**, 521-528 (1991).
15. Mangan, D. F., Welch, G. R. & Wahl, S. M. *J. Immun.* **146**, 1541-1546 (1991).
16. Gaillard, J. L., Berche, P., Mounier, J., Richard, S. & Sansonetti, P. J. *Infect. Immun.* **55**, 2822-2829 (1987).
17. Tilney, L. G. & Portnoy, D. A. *J. Cell Biol.* **109**, 1597-1608 (1989).
18. Waring, P. *J. biol. Chem.* **265**, 14476-14480 (1990).

ACKNOWLEDGEMENTS. We thank J. Hacker for plasmid pANN202-312 and G. Milon for discussion. A.Z. is supported by an EMBO fellowship. This work was supported in part by the Thrasher Research Fund.

Hsp90 chaperones protein folding *in vitro*

Hans Wiech*, Johannes Buchner†‡, Richard Zimmermann* & Ursula Jakob†

* Zentrum Biochemie/Abteilung Biochemie II der Universität Göttingen, Gosslerstrasse 12d, 3400 Göttingen, Germany

† Institut für Biophysik und Physikalische Biochemie, Universität Regensburg, Universitätsstrasse 31, 8400 Regensburg, Germany

THE heat-shock protein Hsp90 is the most abundant constitutively expressed stress protein in the cytosol of eukaryotic cells^{1,2}, where it participates in the maturation of other proteins, modulation of protein activity in the case of hormone-free steroid receptors, and intracellular transport of some newly synthesized kinases³⁻⁵. A feature of all these processes could be their dependence on the formation of protein structure. If Hsp90 is a molecular chaperone involved in maintaining a certain subset of cellular proteins in an inactive form, it should also be able to recognize and bind non-native proteins, thereby influencing their folding to the native state. Here we investigate whether Hsp90 can influence protein folding

in vitro and show that Hsp90 suppresses the formation of protein aggregates by binding to the target proteins at a stoichiometry of one Hsp90 dimer to one or two substrate molecule(s). Furthermore, the yield of correctly folded and functional protein is increased significantly. The action of Hsp90 does not depend on the presence of nucleoside triphosphates, so it may be that Hsp90 uses a novel molecular mechanism to assist protein folding *in vivo*.

We used purified bovine pancreatic Hsp90 to study the refolding of the model protein citrate synthase from the

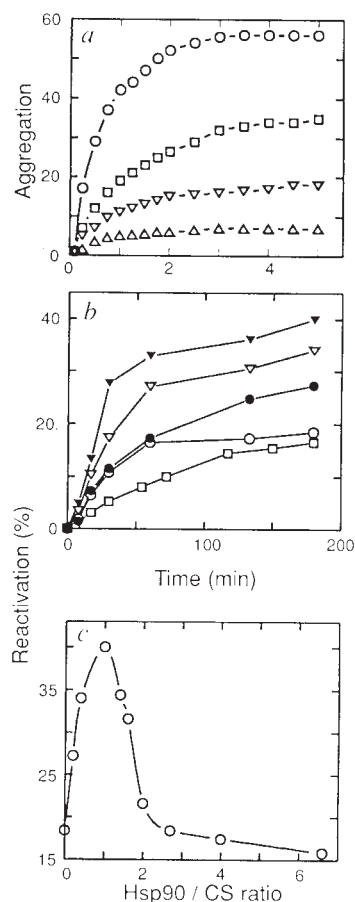
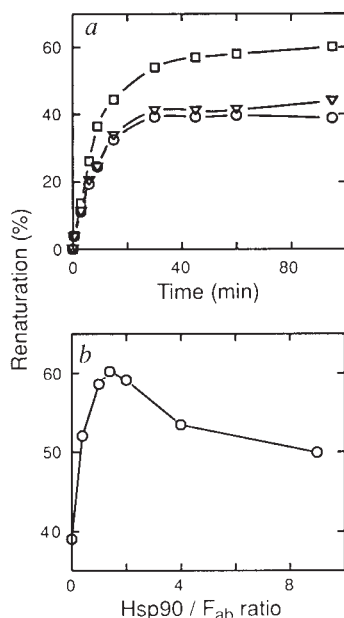


FIG. 1 Hsp90 influences refolding of denatured citrate synthase. *a*, Light scattering (arbitrary units) during the time course of refolding of citrate synthase (0.15 μM) in the presence of (□) 0.30 μM Hsp90, (▽) 0.60 μM Hsp90, (◇) 1.20 μM Hsp90 and (○) in the absence of Hsp90. *b*, Time course of reactivation of citrate synthase in the presence of (●) 0.03 μM Hsp90, (▼) 0.06 μM Hsp90, (■) 0.15 μM Hsp90, (□) 1.0 μM Hsp90 and (○) in the absence of Hsp90 at 20 °C. *c*, Dependence of reactivation of citrate synthase on the molar ratio of Hsp90/citrate synthase. Reactivation yields were determined 3 h after initiation of refolding.

METHODS. Hsp90 was purified to homogeneity from bovine pancreas¹⁰. Hsp90 was predominantly dimeric, as determined by size-exclusion chromatography and non-denaturing gel electrophoresis. Protein concentrations were determined as described¹¹. The concentration of citrate synthase was calculated using an extinction coefficient of 1.78 for a 0.1% solution¹². Concentrations of Hsp90 and citrate synthase given refer to the monomers. Citrate synthase was denatured in 50 mM Tris-HCl, 2 mM EDTA, 6 M guanidine hydrochloride, 20 mM dithiothreitol, pH 8.0, for 1.5 h at 20 °C. For light scattering experiments, concentration of citrate synthase was 30 μM. Renaturation was initiated by a 200-fold dilution in renaturation buffer containing 40 mM HEPES, 20 mM KOH, 50 mM KCl, 10 mM (NH₄)₂SO₄, 2 mM magnesium acetate, 0.5 mM EDTA, pH 7.5 (buffer A). Dilution and subsequent renaturation was in cuvettes with vigorous stirring at 20 °C. Light scattering was measured with a Perkin Elmer MPF2A with excitation and emission at 500 nm. The spectral band width was 2 nm for both excitation and emission. For measuring the restoration of enzymatic activity, denatured citrate synthase (15 μM) was diluted 100-fold in a glass vessel containing buffer A with vigorous stirring. Activity was assayed as described¹³ after various times of refolding at 20 °C. The activity of the refolding synthase is given as the percentage relative to a control sample of native enzyme at 0.15 μM.

‡ To whom correspondence should be addressed.



denatured state. Dimeric mitochondrial citrate synthase refolds spontaneously with low efficiency⁶, most of the protein forming insoluble aggregates under these conditions. It has been shown previously that in the presence of the chaperone GroEL from *Escherichia coli*, aggregation is suppressed and refolding to the functional form occurs in the presence of GroES and ATP⁶. If Hsp90 plays a general role in protein folding, it should prevent the non-native protein from forming large inactive aggregates. Light-scattering experiments showed that in the absence of Hsp90, citrate synthase aggregates rapidly on dilution of the guanidine-denatured protein into refolding buffer (Fig. 1a). But in the presence of Hsp90 aggregation is drastically reduced. At a twofold excess of Hsp90 monomer to citrate synthase, monomer aggregation is suppressed to ~50%. When Hsp90 is present in an eightfold molar excess, aggregation is suppressed completely. Nucleoside triphosphates had no effect on this reaction (data not shown). Suppression of aggregation should result in an increased yield of functional protein. Therefore, we examined the yield of refolded and enzymatically active protein in the absence and presence of Hsp90. In the absence of Hsp90, about 15% of the citrate synthase reached the native state (Fig. 1b); in the presence of Hsp90, the amount of refolded protein increased. At equimolar concentrations of Hsp90 to citrate synthase monomers, reactivation yield was about 40% (Fig. 1b and c). Higher concentrations of Hsp90 gave an increased inhibition of renaturation. Compared with spontaneous reactivation, the kinetics with excess Hsp90 were slow and yields were even lower than in the absence of Hsp90 (Fig. 1c). One explanation for these observations is that citrate synthase released from Hsp90 was rebound by Hsp90 molecules under these conditions.

To test whether the chaperoning of folding is a general function of Hsp90, we investigated the influence of Hsp90 on the refolding of another protein with different secondary structure composition and different folding characteristics. As a model protein we chose the Fab fragment of a monoclonal antibody.

FIG. 2 Hsp90 stimulates folding of a denatured antibody Fab fragment. *a*, Time course of reactivation of the Fab fragment (final concentration, 0.15 μM) in the presence of 0.15 μM Hsp90 (□) or 0.15 μM Hsp90 preincubated at 95 °C for 15 min (▽), and in the absence of Hsp90 (○). Similar effects of Hsp90 were obtained using a Fab fragment that lacked the interchain disulphide bridge (data not shown). *b*, Dependence on the reactivation of the Fab fragment on the ratio of Hsp90/Fab. Reactivation yields were determined 1.5 h after the initiation of refolding.

METHODS. MAK 33, a murine monoclonal antibody directed against human creatine kinase¹⁴ was obtained from Boehringer Mannheim. The Fab fragment was prepared by proteolytic digestion of the intact antibody¹⁵. For Fab, the extinction coefficient for a 0.1% solution was determined as 1.6 from the sequence¹⁶. The Fab fragment was denatured for 3 h in 50 mM Tris, 2 mM EDTA, 6 M guanidine hydrochloride, pH 8.0, at 20 °C. The concentration of the denatured, disulphide-bridged Fab fragment was 15 μM. As for Fig. 1, Hsp90 concentrations refer to the monomer. Renaturation was by a 100-fold dilution of the denatured Fab fragment in 50 mM Tris, 2 mM EDTA, pH 8.0. Reactivation yields were determined by ELISA as described¹⁷. Activity was assayed after different times of refolding. To prevent further folding of the Fab fragment during ELISA, aliquots were preincubated with trypsin (final activity, 300 U per ml) for 10 min at 20 °C. The antigen-binding activity of the refolding Fab is given as the percentage of a control sample of native Fab which was also preincubated with trypsin.

The secondary structure of the Fab fragment consists of β-sheet structures only, whereas citrate synthase is a predominantly helical protein. Unlike the citrate synthase dimer, which is stabilized by only non-covalent interactions, the two different polypeptide chains of the Fab fragment are linked by an inter-chain disulphide bond. In addition there are disulphide bridges in each of the four domains of the antibody fragment. The folding of antibodies and their fragments is well characterized⁷⁻⁹. Refolding of the Fab fragment from the denatured state is a rapid process with a yield for spontaneous reactivation of about 40% under our assay conditions (Fig. 2a). In the presence of Hsp90 monomer at equimolar concentrations, there was a significant increase in yield. Similarly, GroE is also able to facilitate refolding of the Fab fragment, increasing the yield of functional protein to ~70% (ref. 18). Thermally inactivated Hsp90 present during renaturation at the same concentration has no effect on the yield of functional protein (Fig. 2a). Variation of the concentration of Hsp90 during refolding showed that the highest effect is achieved at a 1:1 to 2:1 ratio of Hsp90 monomer to Fab. Although the yield decreased at higher concentrations of Hsp90, there was no inhibition of refolding as described for citrate synthase (Fig. 1c).

In summary, we have shown that Hsp90 is a molecular chaperone stimulating protein folding by preventing non-native proteins from participating in unproductive intermolecular interactions. In this respect Hsp90 is similar to the prokaryotic GroE proteins. Our results are consistent with a model in which one Hsp90 dimer binds one or two molecules of non-native 'substrate protein'. As nucleoside triphosphates have no effect on the function of Hsp90, it may be that in this case the release of the substrate protein is driven by the folding of the non-native protein. Thus Hsp90 and probably its counterparts Grp94 in the endoplasmic reticulum and bacterial HtpG represent a class of molecular chaperones that have a GroE-like effect on protein folding but which function by a completely different mechanism. □

Received 10 March; accepted 19 May 1992.

- Welch, W. J. in *Stress Proteins in Biology and Medicine* (eds Morimoto, R. J., Tissières, A. & Georgopoulos, C.) 223-278 (Cold Spring Harbour, New York, 1990).
- Lindquist, S. C. & Craig, E. A. *Rev. Genet.* **22**, 631-677 (1988).
- Gething, M. J. & Sambrook, J. *Nature* **355**, 33-44 (1992).
- Pelham, H. R. B. *Cell* **46**, 959-961 (1986).
- Ellis, R. J. & Hemmingson, S. M. *Trends biochem. Sci.* **14**, 339-342 (1989).
- Buchner, J. *et al. Biochemistry* **30**, 1586-1591 (1991).
- Goto, Y. & Hamaguchi, K. *J. molec. Biol.* **156**, 891-910 (1982).
- Tsunenaga, M., Goto, Y., Kawata, Y. & Hamaguchi, K. *Biochemistry* **26**, 4367-4373 (1987).
- Kawata, Y. & Hamaguchi, K. *Biochemistry* **30**, 4367-4373 (1991).

- Welch, W. J. & Feramisco, J. R. *J. biol. Chem.* **257**, 14949-14959 (1982).
- Bradford, M. *Analyt. Biochem.* **72**, 248-254 (1976).
- West, S. M. & Price, N. C. *Biochem. J.* **251**, 135-139 (1988).
- Srere, P. A., Brazil, H. & Gonen, L. *Acta chem. scand.* **17**, S129-S134 (1963).
- Buckel, P. *et al. Gene* **51**, 13-19 (1987).
- Johnstone, A. & Thorpe, R. *Immunocytochemistry in Practice* 55-59 (Blackwell Scientific, Oxford, 1987).
- Wetlaufer, D. B. *Adv. Prot. Chem.* **17**, 303-389 (1962).
- Buchner, J. & Rudolph, R. *Bio/Technology* **9**, 157-162 (1991).
- Schmidt, M. & Buchner, J. *J. biol. Chem.* (in the press).

ACKNOWLEDGEMENTS. This work was supported by the Deutsche Forschungsgemeinschaft and by the Fonds der Chemischen Industrie.

Protein images obtained by STM, AFM and TEM

Hideo Arakawa, Kazuo Umemura and Atsushi Ikai

Scanning tunnelling microscopy and atomic force microscopy, one scanning the tunnelling current and the other the repulsive atomic force between sample and probe, can give high-quality surface topographies of proteins, which have been difficult to obtain by more conventional methods such as transmission electron microscopy.

SCANNING tunnelling microscopy (STM) and atomic force microscopy (AFM) are potentially the most powerful techniques for probing biomolecular structures¹⁻⁶, although the quality of the observed images can vary, especially for proteins. To improve the situation of STM and AFM imaging of proteins, it is important to know how the same proteins look under STM and AFM, and to compare the results with those obtained by electron microscopy. Here, we report images of the same protein molecule (turtle α -macroglobulin) obtained by STM, AFM and transmission electron microscopy (TEM) with submolecular resolution. The results show that STM and AFM images of uncoated protein obtained under air have a comparable or better resolution to that of TEM with negative staining. Consequently,

these techniques may be of tremendous potential for obtaining previously unavailable information on the surface topography of a single protein molecule and non-crystallizable biological structures such as cell membranes.

One of the most important features of STM and AFM is their ability to image proteins under physiological conditions. This makes it possible to register changes in conformational structure and subunit rearrangements of proteins in their natural environments. Drake *et al.*⁷ observed in real-time the polymerization of fibrin using AFM, although, unfortunately, the molecular structure of the protein was not clearly resolved in their study. AFM images of fibrinogen have also been obtained with good resolution by Wigren and colleagues⁸.

To apply the real-time observation study to protein chemistry, it is important to establish the quality of protein images obtained with STM and AFM under air or water. Although there have been reports on the imaging of proteins by STM or AFM⁹⁻¹³, comparisons of the techniques using the same protein generally have not been carried out. Therefore, in order to evaluate such images, it was necessary to compare images obtained by different methods.

Images of linear, oligomer turtle α -macroglobulin obtained by STM, AFM and TEM are shown in Fig. 1. For STM and AFM, aliquots of oligomer protein solution were dropped on a cleaved, highly-oriented pyrolytic graphite (HOPG, Union Carbide, Ohio, USA) surface, incubated with

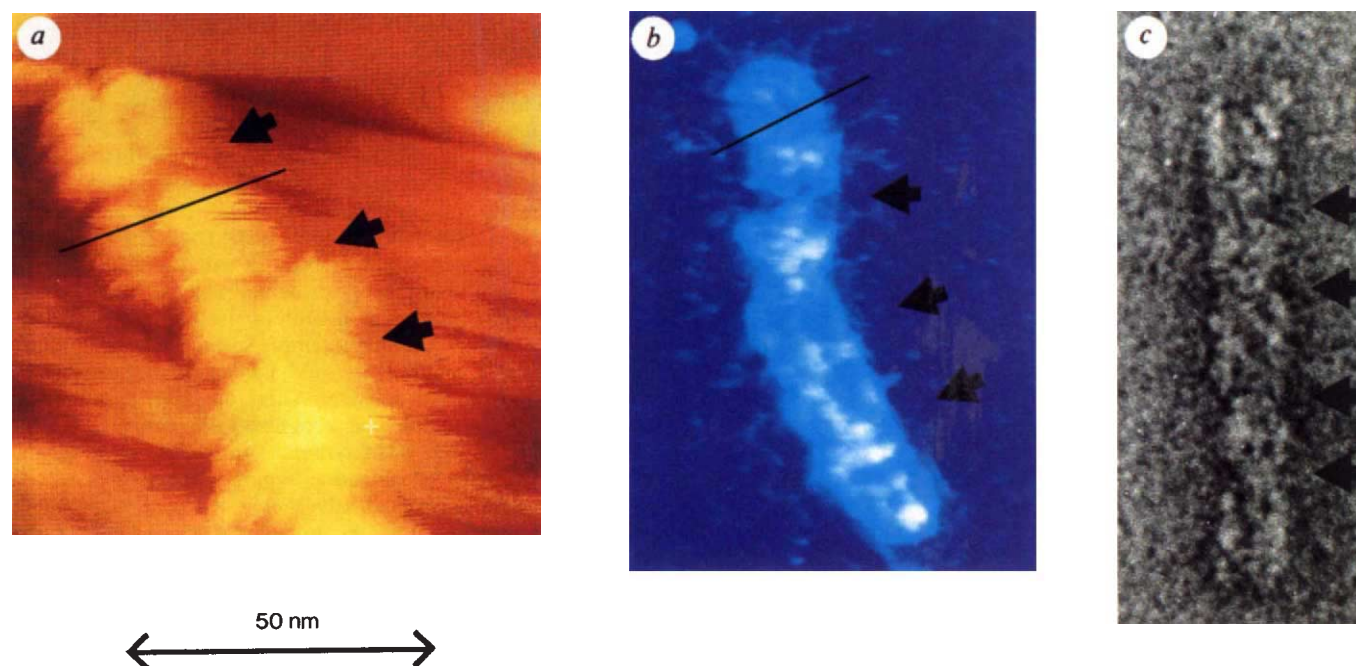


FIG. 1 Oligomer of turtle α -macroglobulin imaged by STM (colour-scale image of α -macroglobulin tetramer) (a), AFM (colour-scale image of α -macroglobulin tetramer) (b), and TEM (negatively-stained image of α -macroglobulin pentamer) (c). Arrows in each figure indicate boundaries of each α -macroglobulin molecule in the oligomers. All three images are the same magnification. **METHODS.** Alpha-macroglobulin oligomer was prepared according to Osada *et al.*¹⁴ and crosslinked with 0.01 % glutaraldehyde to fix the protein in oligomeric forms. After dialysis against 10 mM ammonium sulphate to remove salts and glutaraldehyde, the solution was used for (i) STM and AFM or (ii) TEM. (i) For STM and AFM, glutaraldehyde was added again (final concentration 0.01 %), 50 μ l of the mixture was put on a cleaved, highly-oriented pyrolytic graphite (HOPG) and incubated for 30 min at 37 °C. The surface of the graphite was rinsed with distilled water and then dried in a vacuum jar by aspirator for about 1 h at room temperature. STM was performed in constant-current mode, setpoint 1.0 nA, bias 200 mV (tip negative), scan rate 1.9 Hz, scan range 100 nm $\times$ 100 nm using a NanoScope II instrument with Pt:Ir (8:2) Nanotip (Digital Instruments, California, USA). AFM was performed with set point 0.27 nN, scan speed 1.0 μ m s⁻¹, scan range 100 nm $\times$ 100 nm using a SFA300 instrument (Seiko Instruments Ltd, Tokyo, Japan) with Microlever (Park Scientific, California, USA). (ii) For TEM, the protein was observed on an H-7000 electron microscope (Hitachi Ltd, Tokyo, Japan) after negative staining with 3 % uranyl acetate on carbon-coated Formvar-supported film (Monsanto Company, Missouri, USA) as described in ref. 14. The TEM images obtained were the same as those images obtained without prefixation with glutaraldehyde, as reported by Osada *et al.*¹⁴.

ADVERTISEMENTS

\$3.50/Base

**Custom DNA
Synthesis**

No Set-up Fee,
Shipped in 24 hours.

**1-800-788-4362
1-800-786-4362 Fax**

**KEYSTONE
Laboratories, Inc.**

Reader Service No. 33

**Highest Quality
Oligonucleotides**

Synthesised, purified and shipped in 3 days

£2.75 per base

(All inclusive)

0.2 µM scale (approx. 500 µg)

- Antisense (completely or partially sulphurised)
- Extensive range of labelling and modifications
- Full design and analysis facilities
- Gene cassettes, bulk order and contract discounts

Molecular Medicine Unit

King's College School of Medicine & Dentistry

Tel. 071 - 326 3126 FAX 071 - 733 3877

Reader Service No. 56

Synthetic Oligonucleotides

Rapid Quality Worldwide Service

£2.30 per base, no setup!

(0.2 µmole - limited offer)

option of H.P.L.C. purification - £30
biotinylation, fluorescent labelling
phosphothioates, many other modifications

OSWEL DNA SERVICE

Department of Chemistry, University of Edinburgh
Kings Buildings, West Mains Road,
EDINBURGH EH9 3JJ

Tel: 031 650 4792 Fax: 031 662 4054

Reader Service No. 9

PACEMAKERSM

**The multiple peptide
synthesis and conjugation
service**

- High purity (up to 95%).
- Quantities for a range of uses.
- Single or multiple peptides.
- HPLC trace with each peptide.
- Terminal modification at no extra charge.
- Conjugation to carrier proteins.

from **£80 per peptide**

AFFINITY Research Products Ltd.

GPT Business Park, Technology Drive,
Nottingham, NG9 2ND, UK.

Tel: +44/0 602 436100; Fax: +44/0 602 436300

Reader Service No. 23

glutaraldehyde, and then rinsed with distilled water and dried. For TEM, the protein was negatively stained on carbon-coated film using the methods described by Osada *et al.*¹⁴. In the STM image (Fig. 1a), four α-macroglobulin molecules linked in a head-to-tail manner were identified. A groove running in the centre parallel to the longer axis was apparent in the first two molecules. The groove is a characteristic submolecular structure of macroglobulins¹⁵. The protein mass on each side of the groove is divided into three globular domains by shallow indentations. Each molecule had a dimension of about 22 nm × 22 nm.

In the AFM image of Fig. 1b, a tetramer of α-macroglobulin linked in a head-to-tail manner was observed. Although the centre groove was not clear in Fig. 1b, the image does show finer resolution of the surface topography of the protein than the image in Fig. 1a. The AFM image in Fig. 1b reproduced the major characteristics of the TEM image in Fig. 1, such as the open disposition of string-like subunits within a molecule, giving confidence to the identity of the AFM image. The existence of the centre groove was implied by taking a cross-section of the AFM image (see Fig. 2b). The dimension of each molecule in the AFM image was about 16 nm × 22 nm — a little narrower in width than the STM image. The resolution of STM and AFM is limited by the curvature of the probe. For mechanical reasons, AFM probes presently have larger curvatures than those for STM. The curvature of the AFM probe was estimated to be about 20 nm or greater but the resolution of the resulting AFM images of sodium phosphate crystals was about 2–3 nm — a much better result than expected from the tip curvature. Sodium phosphate buffer dried on HOPG showed a characteristic crystal structure under AFM and grains of 2 nm in diameter are clearly observed with good reproducibility.

The TEM image (Fig. 1c) shows a pentamer of α-macroglobulin, which confirmed results reported earlier by Osada *et al.*¹⁴. In this image, each molecule had a dimension of about 14 nm × 18 nm, which was still smaller than the molecules in the AFM image. As the sample for TEM was negatively stained, the image shows the areas where heavy metal was excluded. In the TEM experiment, penetration of staining material into finer molecular crevices and accidental covering of the sample by heavy metals may produce smaller images of proteins than their actual size. We are thus convinced that the STM and AFM images obtained above are those of turtle α-macroglobulin. Control experiments were carried out on empty HOPG as various artifacts have been reported using HOPG^{16,17}.

Cross-section profiles

Cross-sections of molecules from STM and AFM images are shown in Fig. 2. The height of the molecule was 3.4 nm and 6.1

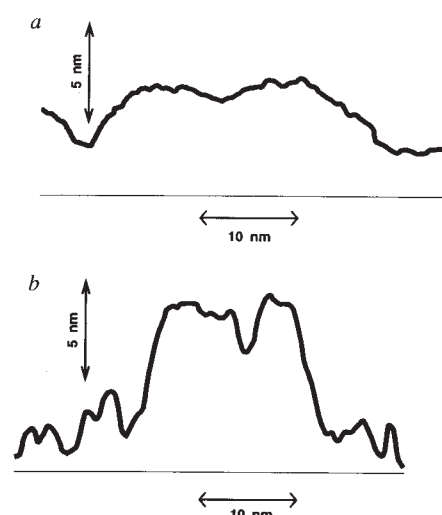


FIG. 2 Cross-sections of α-macroglobulin obtained by STM (a) and AFM (b). (The cross-sections are taken at points indicated by solid lines on Figs 1a and 1b.) Both figures are expanded twice in the vertical direction.

nm, as measured by STM and AFM respectively. The depth of the central groove was 0.9 nm and 2.4 nm, respectively. There were some characteristic differences between the cross-sections obtained by STM and AFM and these are thought to be due primarily to the different shapes of tips used on the respective instruments and on the different principles on which the instruments were built. Scanning tunnelling microscopy produced wider profiles of α-macroglobulin than AFM, whereas AFM produced taller images. The difference in electronic surface properties between protein and graphite may be one of the reasons why the image obtained by STM was lower. A small amount of salt should be on the surface of proteins, making AFM images taller than protein itself. STM works by scanning a sample surface keeping the tunnelling barrier between the probe and sample constant. Therefore, salt, an insulator, should be recognized as a part of the barrier in STM, whereas, in AFM, it should be recognized as a part of the sample. When residual salts were not washed out before drying, the protein was imaged much taller by AFM.

Further research is required to clarify the mechanism that underlies the difference between STM and AFM images of the same protein. Application of scanning tunnelling spectroscopy, a variant of STM, will be helpful in determining further information on the electronic surface properties. If it becomes possible to measure both atomic force and tunnelling current simultaneously with one probe on proteins, it should be possible to interpret the differences in images obtained by STM and AFM more directly based on their different working principles.

In conclusion, the ability of STM and AFM to produce comparable images to

TEM was shown. Further, we have illustrated several advantages to these newer techniques. STM and AFM make an image of the surface topography, whereas TEM makes a projection of the sample. It is this capability that makes it possible to distinguish the groove on the surface of the α -macroglobulin as being on top of the molecule and not inside or under the molecule (this is not possible using TEM with negative staining). In experiments using STM and AFM, the sample is free from electron beam damage, which can be severe with TEM. Additionally, using these newer techniques, images can be obtained in air with good resolution, as shown here. This opens up the possibility of observing native and functional protein under physiological conditions, which, so far, has been impossible. Although STM and AFM are proving to be powerful techniques for imaging surface topography, they are not as good as TEM at drawing the boundary of a molecule due to the effects of the shape of the probe. Therefore, complementary use of the newer techniques with more conventional methods is desirable. □

Hideo Arakawa and Kazuo Umemura should be considered equally as first author. H. A., K. U. and Atsushi Ikai are at the Faculty of Bioscience and Biotechnology, Tokyo Institute of Technology, Nagatsuta, Midori-ku, Yokohama 227, Japan. The authors thank T. Akaike (Kanagawa Academy of Science and Technology) for use of the STM and M. Shishido (Tokyo Institute of Technology) for use of the AFM. The authors also thank T. Osada and M. Nishigai. This work was supported by the Japanese Ministry of Education and by the Nissan Science Foundation. For more information, fill in reader service number 100.

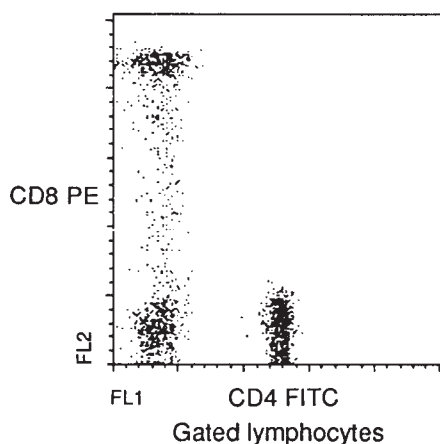
- Hoh, J. H., Lal, R., John, S. A., Revel, J. P. & Arnsdorf, M. F. *Science* **253**, 1405-1408 (1991).
- Driscoll, R. J., Youngquist, M. G. & Baldeschwieler, J. D. *Nature* **346**, 294-296 (1990).
- Lindsay, S. M., Thundat, T., Nagahara, L., Knipping, U. & Rill, R. L. *Science* **244**, 1063-1064 (1989).
- Dunlap, D. D. & Bustamante, C. *Nature* **342**, 204-206 (1989).
- Arscott, P. G., Lee, G., Bloomfield, V. A. & Evans, D. F. *Nature* **339**, 484-486 (1989).
- Hörber, J. K. H., Lang, C. A., Hänsch, T. W., Heckl, W. M. & Möhwald, H. *Chem. Phys. Lett.* **145**, 151-158 (1988).
- Drake, B. et al. *Science* **243**, 1586-1589 (1989).
- Wigren, R., Elwing, H., Erlandsson, R., Wellin, S. & Lundström, I. *FEBS Lett.* **280**, 225-228 (1991).
- Leatherbarrow, R. J., Stedman, M. & Wells T. N. C. *J. molec. Biol.* **221**, 361-365 (1991).
- Verneti, L. A., Nowlin, C. L. A., Hameroff, S. R., Gandolfi, A. J., Lee, Y. C. & Sarid, D. *J. Vac. Sci. Technol.* **B9**, 1223-1226 (1991).
- Weisenhorn, A. L. *Biophys. J.* **58**, 1251-1258 (1990).
- Edstrom, R. D., Meinke, M. H., Yang, X., Yang, R., Ellings, V. & Evans, D. F. *Biophys. J.* **58**, 1437-1448 (1990).
- Edstrom, R. D., Meinke, M. H., Yang, X., Yang, R. & Evans, D. F. *Biochemistry* **28**, 4939-4942 (1989).
- Osada, T., Nishigai, M., & Ikai, A. *J. Ultrastruct. molec. Struct. Res.* **101**, 62-74 (1988).
- Ikai, A., Nishigai, M., Osada, T., Arakawa, H. & Kikuchi, M. *J. Protein Chem.* **6**, 81-93 (1987).
- Clemmer, C. R. & Beebe, T. P. Jr *Science* **251**, 640-642 (1991).
- Salmeron, M., Beebe, T., Odriozola, J., Wilson, T., Ogletree, D. F. & Siekhaus, W. J. *Vac. Sci. Technol.* **A8**, 635-641 (1990).

Detection limits

New products featured this week include a gel designed specifically for the detection of point mutations and a nonradioactive assay for detecting HIV DNA.

HYDROLINK MDE (mutation detection enhancement) gel from AT Biochem is designed specifically for the **detection of point mutations** and eliminates the need for denaturing gradient methods or special apparatus (*Reader Service No. 101*). The polymerase chain reaction products of normal and mutant alleles can be separated using nondenaturing electrophoresis in MDE gel. Resolution is based on conformational differences in either double-stranded molecules (heteroduplex analysis) or single-stranded DNA (single-strand conformational polymorphism analysis, SSCP). AT Biochem says MDE gel will resolve heteroduplex DNA fragments up to 900 base pairs in length differing by a single base pair. Two kits are available: the \$250 (US) kit for heteroduplex analysis and the \$195 (US) kit for SSCP analysis.

Exalpha has developed a new line of two-colour products for the **detection of cell-surface antigens** in flow cytometry and fluorescent microscopy (*Reader Service No. 102*). The Bi-Test line of reagents can be used for the study of AIDS virus



Peripheral blood lymphocytes stained with Exalpha's Bi-Test CD4 FITC-CD8 PE two-colour staining kit.

infection, T- and B-cell subsets, NK cells, T-cell activation, cellular adhesion and many other applications. The reagents included in the Bi-Test line are: CD4 FITC-CD8 PE; CD8 PE-CD4 FITC; CD5 FITC-CD19 PE; HLA-DR FITC-CD4 PE and the corresponding controls IgG1 FITC-IgG1 PE and IgG2A FITC-IgG1 PE.

Last May, Cambridge Research Biochemicals launched SPOTs, a novel system for **mapping antibody epitopes** (*Reader Service No. 103*). SPOTs uses a cellulose membrane upon which peptides may be synthesized. Each SPOTs membrane

allows 96 peptides to be made in an 8 × 12 format. To scan antibodies for epitopes, a complete set of overlapping peptides is generated from the antigen using the IBM software provided. CRB says synthesis is rapid and simple, taking only 90 minutes per amino acid. Once synthesized, the SPOTs membranes are incubated with antibody — just like with a dot blot. The SPOTs that have bound antibody turn blue after incubation with substrate, thus indicating the antibody's epitope. The membranes can then be regenerated and probed again with antibody.

Seradyn announces two new **dyed microparticles**, Color-Rich (dyed) and Fluoro-Max (fluorescent) microparticles, which are monodisperse particles prepared by proprietary emulsion polymerization techniques that incorporate the dyes internally. (*Reader Service No. 104*). Internal dyeing of the particles is designed to ensure maximum colour saturation, prevent dye leaching in aqueous media and leave the surface free for a variety of immunological applications. Even at low concentrations, the Color-Rich and Fluoro-Max microparticles exhibit deep colour and brilliant fluorescence, Seradyn says. The Fluoro-Max microparticles can be illuminated and detected with epifluorescence, fluorescent microscopes, fluorescence spectrophotometers, fluorescent cell sorters or even back lights. Stated applications include: clinical diagnostic assays, membrane pore size determinations, flow or fluid mechanism studies, biological transport studies, flow cytometry, flow visualization, filtration media analysis, particle size correlation, immuno/histology studies, cell-surface antigen markers, phagocytosis research and membrane-based diagnostics. The Color-Rich and Fluoro-Max particles are generally available in 0.1-, 0.2-, 0.3- and 0.8- μ m diameters in plain polystyrene and carboxylate-modified surface modifications. Premium-grade particles are available in 0.05 to 0.1 μ m.

Assorted assays

The ENZO Microplate Assay for HIV DNA is a **nonradioactive kit for detecting HIV DNA** in a microtitre well, which should find application in the areas of epidemiology, diagnosis and treatment (*Reader Service No. 105*). HIV DNA can be assayed directly if there is sufficient target DNA present, or it can be assayed in procedures employing target amplification. The kit provides materials for the colorimetric detection of HIV proviral DNA or viral RNA that has been reverse transcribed. Enzo says direct and

quantitative HIV detection offers a distinct advantage over serological assays: the identification of proviral DNA can be made independent of the presence or absence of antibodies, which appear several months after initial infection. The assay is based on a two-probe hybridization method and can be carried out entirely in a 96-well microtitre plate or, alternatively, microwell strips. A positive reaction is indicated by the generation of colour, which can be measured by a microplate reader. Using this format, Enzo says that fewer than 10 proviral sequences can be detected.

A leukotriene $C_4/D_4/E_4/F_4$ competitive enzyme immunoassay kit for the quantitative determination of human leukotriene $C_4/D_4/E_4/F_4$ levels in biological fluids can now be ordered from Advanced Magnetics (Reader Service No. 106). The assay is based on the variable amounts of leukotriene C_4 in the samples with a fixed amount of alkaline phosphate-labelled leukotriene C_4 for a limited number of sites on the monoclonal anti-leukotriene $C_4/D_4/E_4/F_4$ antibody. Precision-coated wells are supplied with each kit to provide both consistency and flexibility. Results can be obtained within 24 hours and the sensitivity of detection is stated to be 27.7 pg ml^{-1} . Each assay kit contains sufficient reagents for 96 tests, including a standard curve.

Boehringer has introduced a human tumour necrosis factor- α (h-TNF- α) ELISA kit for the quantitative determination of TNF- α in the physiological and pathologically increased range ($5\text{--}1,000 \text{ pg ml}^{-1}$) (Reader Service No. 107). The two mouse monoclonal antibodies used are directed against two different epitopes of TNF- α : one of which recognises the biologically effective epitope. This, Boehringer says, leads to the determination of only bioactive TNF- α in this assay system. There are three steps to the assay. During the first incubation step, TNF- α of samples/standards is bound simultaneously by biotinylated antibody and peroxidase-conjugated detection antibody (one-step immunoreaction). The complex binds tightly to the streptavidin-coated microtitre plate. Second, the plate is washed. And, third, the peroxidase substrate tetramethylbenzidine is added. The colour that develops is proportional to the concentration of TNF- α . Boehringer says that the sensitivity of the assay is approximately 10 pg ml^{-1} .

IBL has introduced an ELISA for the quantitative determination of soluble CD14 in human serum, plasma, cell culture supernatants and other biological fluids (Reader Service No. 108). Evidence is available of a role for soluble CD14 in polytrauma, sepsis and inflammation, where it may be a useful prognostic marker. The assay, which involves two one-hour incubations, uses a convenient 12×8

microtitre plate format with wells precoated with a specific monoclonal antibody. IBL says that the detection limit of the assay is 1 ng ml^{-1} .

■ Image makers

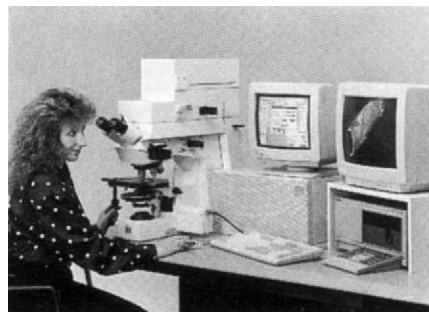
Gels, dot blots, slot blots or any radioactive sample can be imaged without the need for X-ray film, intensifiers, or low-temperature film storage and exposure with the new Model GS-250 **molecular imager system** from Bio-Rad (Reader Service No. 109). To use the system, the researcher exposes the radioactive sample to erasable and reusable storage phosphor screens, which, according to Bio-Rad, are sensitive to not only ^{32}P , but also ^{35}S , ^{14}C , ^{125}I , X-rays and others. In



Bio-Rad's phosphor image analysis system.

minutes, a latent image is generated, which is then scanned by the patented fibre-optic laser scanner. Data from one- and two-dimensional gels, protein gels or blots are then displayed, analysed and output. Software packages are available for both Windows-compatible and Macintosh-compatible computers, which allow quick quantification of DNA, RNA or protein samples; molecular weight analysis; plotting of one-dimensional density profiles or histograms; and a myriad of image viewing and processing manipulations.

The LSM310 **laser scanning microscope** newly introduced by Carl Zeiss is a third-generation instrument that features a high-performance computer and Windows interface (Reader Service No. 110). Central to the LSM system is a research microscope with integrated equipment for confocal laser scanning. Depending on the fluorescence specimen used, lasers with the wavelengths 365 nm (UV), 488 nm, 514 nm, 543 nm, 633 nm or any other wavelength can be used. Up to three lasers can be attached to cater for multiple staining. The acquisition rate of images of $512 \times 512 \text{ pixels} \times 8\text{-bit}$



LSM310 confocal laser scan microscope.

grey levels is two images per second, Zeiss says. With the ICS (infinity colour-corrected system) of optics from Zeiss, users have more than one hundred objectives to choose from. Two photomultipliers for the recording of confocal fluorescence, and one detector for transmitted light can be activated simultaneously. This, Zeiss says, permits the "live" recording of the coloured overlay of three complete images of $3 \times 8\text{-bit RGB}$. The incorporation of a computer allows the user to connect the LSM310 to local networks and to use evaluation software for kinetic analyses.

MP4+ is Polaroid's latest **multipurpose camera system** (Reader Service No. 111). For electrophoresis gel recording applications, there is a sub-illuminating white-light baseboard and a filter kit for the four most common staining techniques. Colour photography applications can now benefit from quartz-halogen lighting, which matches perfectly with tungsten films, Polaroid says. An electronic flash option is designed to eliminate complicated colour filtration and repeat shots due to "missed" balance. For high-magnification macro work, a fibre-optic macro lighting system provides intense focused lighting on small areas of subjects. Like its predecessor, the MP-4, the MP4+ is available with an assortment of pack, sheet and slide film holders.

■ Monoclonals

In addition to androgen, oestrogen and progesterone receptor antibodies, a **monoclonal antibody to glucocorticoid receptor**, BuGR2, is now available from Affinity BioReagents (Reader Service No. 112). The company also offers three polyclonal anti-human glucocorticoid receptor antibodies, each reacting with a different epitope of the human glucocorticoid receptor. Completing the line-up is a polyclonal anti-mineralocorticoid receptor antibody, which can be used to detect both the activated and unactivated form of the receptor in western blots, immunocytochemical and immunoprecipitation studies.

The gut hormone/neuropeptide cholecystokinin has been associated with several normal physiological functions and disorders, including satiety. An **antibody specific for the sulphated form of cholecystokinin octapeptide (CCK-8)** is now available from Affiniti Research Products for immunometric applications, particularly radioimmunoassays (Reader Service No. 113). The antibody does not crossreact with the related gastrins or desulphated CCK-8. It was raised to a modified CCK-8 peptide conjugated to bovine serum albumin. The assay sensitivity is $0.5 \text{ fmol per tube}$ (2 pmol l^{-1} plasma), Affiniti says. The antibody is offered in four sizes (250, 500, 1,000 and 5,000 tube equivalents), and the peptide is supplied in lyophilized form ready for iodination. □